

# Stability of the shadow projection and the left-curtain coupling<sup>1</sup>

Nicolas Juillet<sup>2</sup>

Institut de Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS, 7 rue René Descartes, 67000 Strasbourg, France.  
E-mail: [nicolas.juillet@math.unistra.fr](mailto:nicolas.juillet@math.unistra.fr)

Received 31 August 2014; revised 16 May 2015; accepted 9 July 2015

**Abstract.** The (left-)curtain coupling, introduced by Beiglböck and the author is an extreme point of the set of “martingale” couplings between two real probability measures in convex order. It enjoys remarkable properties with respect to order relations and a minimisation problem inspired by the theory of optimal transport. An explicit representation and a number of further noteworthy attributes have recently been established by Henry-Labordère and Touzi. In the present paper we prove that the curtain coupling depends continuously on the prescribed marginals and quantify this with Lipschitz estimates.

**Résumé.** Le couplage rideau (gauche), introduit par Beiglböck et Juillet est un point extrémal de l'ensemble des couplages « martingale » entre deux mesures de probabilité réelles prises dans l'ordre convexe. Ce couplage possède des propriétés remarquables quant aux relations d'ordre entre mesures d'une part, et par rapport à un problème de minimisation issu de la théorie du transport optimal d'autre part. Une représentation explicite a été récemment mise en évidence par Henry-Labordère et Touzi. Nous démontrons dans cet article que le couplage rideau dépend continument des mesures marginales prescrites et nous quantifions cette dépendance à l'aide d'inégalités lipschitziennes.

MSC: 60G42; 28A33; 60B10

Keywords: Couplings; Martingales; Convex order; Optimal transport

## 0. Introduction

There are at least two standard methods to couple real random variables, that is to obtain a joint law. The first one is the product (or independent coupling)  $(\mu, \nu) \mapsto \mu \otimes \nu$ , and the other is the quantile coupling  $(\mu, \nu) \mapsto \text{Law}(G_\mu, G_\nu)$  where  $G_\mu, G_\nu$  are the generalised inverse of the cumulative functions of  $\mu, \nu$  (also called quantile functions, see Section 1.2). One can easily convince oneself that both operators are continuous in the weak topology. In this paper we are interested in the continuity of another method, namely the left-curtain coupling  $\pi_{lc} = \text{Curt}(\mu, \nu)$ . It was recently introduced in [2] by Beiglböck and the author<sup>3</sup> and further studied by Henry-Labordère and Touzi [5]. As defined in [2],  $\pi_{lc}$  is the measure with marginals  $\mu$  and  $\nu$  such that for every  $x \in \mathbb{R}$ , the two marginals of  $\pi_{lc}|_{]-\infty, x] \times \mathbb{R}}$  are  $\mu|_{]-\infty, x]}$  and the so-called *shadow* (see Definition 2.1) of the latter measure in  $\nu$ . We advocated that under the additional constraint  $\mathbb{E}(Y|X) = X$  on  $\text{Law}(X, Y)$  (that can be satisfied neither by  $\text{Law}(X) \otimes \text{Law}(Y)$  nor by the quantile coupling, except in degenerate cases),  $\pi_{lc}$  can be considered as the most natural coupling of  $\mu = \text{Law}(X)$

<sup>1</sup>The preprint version of our paper has reference [arXiv:1407.8009](https://arxiv.org/abs/1407.8009). The last part of this preprint will appear separately as “Martingales associated to peacocks by using the curtain coupling.”

<sup>2</sup>The author is partially supported by the “Programme ANR ProbaGeo” (ANR-09-BLAN-0364) and the “Programme ANR JCJC GMT” (ANR 2011 JS01 011 01).

<sup>3</sup>Right-curtain couplings can be defined symmetrically and the corresponding result can be deduced easily. In this paper *curtain coupling* and *monotone coupling* indicate left-curtain couplings and left-monotone couplings respectively.

and  $\nu = \text{Law}(Y)$ . Indeed it is distinct from the quantile coupling but can be considered as its natural counterpart under the martingale constraint. Moreover it enjoys remarkable optimality properties with respect to the martingale variant of the usual transport problem on  $\mathbb{R}$ : the martingale transport problem which was introduced in the context of mathematical finance in [1,4,7]. See Proposition 2.8 for more details on  $\pi_{1c}$ . Our main result is that the operator  $\text{Curt} : (\mu, \nu) \mapsto \pi_{1c}$  is continuous.

**Main theorem.** *The mapping  $\text{Curt}$  is continuous on its domain (Theorem 2.16). More precisely if all the measures have the same mass, then*

$$Z(\text{Curt}(\mu, \nu), \text{Curt}(\mu', \nu')) \leq W^{\mathbb{R}^1}(\mu, \mu') + 2W^{\mathbb{R}^1}(\nu, \nu'),$$

where  $W^{\mathbb{R}^d}$  is the Kantorovich distance of measures in  $\mathbb{R}^d$  (see the beginning of Section 1) and  $Z$  the semimetric defined in Section 2.3.2 (Corollary 2.32 of Theorem 2.31).

Moreover a similar statement with  $W^{\mathbb{R}^2}$  in place of  $Z$  fails in the sense that there exist sequences  $(\mu_n)_n, (\mu'_n)_n$  and  $(\nu_n)_n, (\nu'_n)_n$  such that

$$\lim_{n \rightarrow \infty} W^{\mathbb{R}^2}(\text{Curt}(\mu_n, \nu_n), \text{Curt}(\mu'_n, \nu'_n)) / \max(W^{\mathbb{R}^1}(\mu_n, \mu'_n), W^{\mathbb{R}^1}(\nu_n, \nu'_n)) = +\infty$$

(Example 2.37).

The paper is organised in two sections. In the first part we introduce several notions related to positive measures on  $\mathbb{R}$  that will be used in the second part. In particular we introduce the convex order  $\leq_C$ , the stochastic order  $\leq_{\text{sto}}$  and the usual order  $\leq_+$ . The domain of  $\text{Curt}$  is described by Strassen’s theorem [10] in terms of  $\leq_C$ : it is the set of pairs  $(\mu, \nu)$  satisfying  $\mu \leq_C \nu$ . Hence in terms of mass transportation,  $\mu \leq_C \nu$  denotes a dilation; for every  $x \in \mathbb{R}$ , the mass  $\mu(dx)$  at point  $x$  can be spread in both directions in such a way that for any  $x$  the barycenter of the mass transported from  $x$  is still  $x$ . The distribution after the dilation is  $\nu$ . In the same spirit let us see how one can interpret the other orders: the relation  $\mu \leq_{\text{sto}} \nu$  means that part of the mass of  $\mu$  can be transported in the direction of  $+\infty$  in order to build  $\nu$ . The relation  $\mu \leq_+ \nu$  means that some mass is created, i.e.  $\nu = \mu + \mu'$  where  $\mu'$  is a positive measure. We investigate the interplay between these three orders and four other more or less classical orders, defined in Definition 1.1, on the space of positive measures with finite first moments: the extended order  $\leq_{C,+}$  (that is the same as  $\leq_{+,C}$ ), the orders  $\leq_{+, \text{sto}}$  (or  $\leq_{\text{sto}, +}$ ) and  $\leq_{C, \text{sto}}$  (or  $\leq_{\text{sto}, C}$ ) and finally  $\leq_{C,+, \text{sto}}$  (for which the three indices can also be permuted). We reproduce Theorem 1.7 that may be considered the main result of Section 1.

**Theorem.** *For any sequence  $(\mu_i)_{i=0, \dots, n}$  (with  $n = 2$  or  $3$ ) satisfying the relations  $\mu_{i-1} \leq_{r_i} \mu_i$  for  $i = 1, \dots, n$  one has  $\mu_0 \leq_{r_1, \dots, r_n} \mu_n$ .*

*Conversely if  $\mu_0 \leq_{r_1, \dots, r_n} \mu_n$  one can find a sequence  $(\mu_i)_{i=0, \dots, n}$  such that  $\mu_{i-1} \leq_{r_i} \mu_i$  for every  $i \geq 1$ .*

We start Section 2 with the definitions of the shadow projection (Definition 2.1) and the left-curtain coupling (Definition 2.6), and recall their main properties. We have already stated in Main Theorem a compilation of the results of this section. Let us insist on two ingredients.

The key theorem of this part is Theorem 2.31. It states that the shadow projection  $(\mu, \nu) \mapsto S^\nu(\mu)$  is a Lipschitzian map for the Kantorovich metric  $W^{\mathbb{R}^1}$ . Corollary 2.32 is only a reformulation of this result for the left-curtain coupling  $\text{Curt}$ . One important preliminary result for Theorem 2.31 is Lemma 2.33, on the monotonicity of the shadow projection with respect to the stochastic order.

The other ingredient is a new modified support  $\text{spt}^*(\pi)$ , which might be the most important definition of this paper. It leads to a second proof of the continuity of  $\text{Curt}$ . For a positive measure  $\pi$  on  $\mathbb{R}^2$ , we introduce it in the following way: let  $A$  be the set of  $x \in \mathbb{R}$  such that  $\pi([-\infty, x[ \times ]x, \infty]) = 0$ , and  $A^- \subseteq A$  the subset of points that are isolated on the right in  $A$ . We set

$$\text{spt}^*(\pi) = (\text{spt}(\pi) \setminus (A^- \times \mathbb{R})) \cup \bigcup_{\mu(x) > 0} \{x\} \times \text{spt} \pi_x \subseteq \mathbb{R}^2,$$

where  $\mu$  is the first marginal of  $\pi$  and  $(\pi_x)_{x \in \mathbb{R}}$  is some  $\mu$ -almost surely unique family of conditional probability laws of  $\pi$  (notice that  $\text{spt}^*(\pi)$  does not depend on the particular choice). For instance  $\pi = 1/2(\delta_0 + \lambda_{[0,1]})^2$  has reduced support  $\{(0, 0)\} \cup (]0, 1[ \times ]0, 1[)$  and support  $[0, 1]^2$ . The reduced support simplifies the definition of *left-monotone couplings*. According to a result in [2] (Proposition 2.8 in the present paper) this is one of three equivalent properties for couplings: optimal, left-monotone or left-curtain. With Proposition 2.14, it becomes possible to determine whether a coupling  $\pi$  is left-monotone only by testing the triples of points in  $(\text{spt}^*(\pi))^3$ , while the original definition involves *some undefined*  $\Gamma \subseteq \mathbb{R}^2$  of full  $\pi$ -measure. In particular, as explained in Example 2.11, the property to be checked for left-monotonicity may fail for triples of  $\Gamma = \text{spt}(\pi)$  but hold for  $\Gamma = \text{spt}^*(\pi)$ . The introduction of  $\text{spt}^*$  allows to extrapolate the usual optimal transport proof of the continuity of monotone couplings [13, Theorem 5.20] (Theorem 2.16).

### 1. Reminders about the stochastic and convex orders

We consider the space  $\mathcal{M}$  of positive measures on  $\mathbb{R}$  with finite first moments. The subspace of probability measures with finite expectations is denoted by  $\mathcal{P}$ . For  $\mu, \nu \in \mathcal{M}$ , the Kantorovich distance defined by

$$W(\mu, \nu) = \sup_{f \in \text{Lip}(1)} \left| \int f \, d\mu - \int f \, d\nu \right| \tag{1}$$

endows  $(\mathcal{P}, W)$  with  $\mathcal{T}_1$ , the usual topology for probability measures with finite first moments. In the definition, the supremum is taken among all 1-Lipschitzian functions  $f : \mathbb{R} \rightarrow \mathbb{R}$ . We also consider  $W$  with the same definition on the subspace  $m\mathcal{P} = \{\mu \in \mathcal{M} \mid \mu(\mathbb{R}) = m\} \subseteq \mathcal{M}$  of measures of mass  $m$ .

According to the Kantorovich duality theorem, an alternative definition in the case  $\mu, \nu \in \mathcal{P}$  is

$$\inf_{(\Omega, \mathcal{F}, \mathbb{P})} \mathbb{E}(|Y - X|), \tag{2}$$

where  $X, Y : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$  are random variables with marginals  $\mu$  and  $\nu$ . The infimum is taken among all joint laws  $(X, Y)$ , the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  being part of the minimisation problem. Note that without loss of generality  $(\Omega, \mathcal{F}, \mathbb{P})$  can be assumed to be  $([0, 1], \mathcal{B}, \lambda)$  where  $\lambda$  is the Lebesgue measure and  $\mathcal{B}$  the  $\sigma$ -algebra of Borel sets on  $[0, 1]$ .

A special choice of 1-Lipschitzian function is the function  $f_t : x \in \mathbb{R} \rightarrow |x - t| \in \mathbb{R}$ . Therefore if  $\mu_n \rightarrow \mu$  in  $\mathcal{M}$ , the sequence of functions  $u_{\mu_n} : t \mapsto \int f_t(x) \, d\mu_n(x)$  pointwise converges to  $u_\mu$ . The converse statement also holds if all the measures have the same mass and barycenter (see [2, Proposition 4.2] or directly [6, Proposition 2.3]). For every  $\mu \in \mathcal{M}$ , the function  $u_\mu$  is usually called the *potential function* of  $\mu$ .

A measure  $\pi$  on  $\mathbb{R}^2$  with marginals  $\mu$  and  $\nu$  is called a *transport plan* from  $\mu$  to  $\mu$  or a *coupling* of  $\mu$  and  $\nu$ . Let  $\Pi(\mu, \nu)$  be the space of transport plans of mass 1. The subspace  $\Pi_M(\mu, \nu)$  is defined as follows

$$\Pi_M(\mu, \nu) = \{ \pi = \text{Law}(X, Y) \in \Pi(\mu, \nu), \mathbb{E}(Y|X) = X \},$$

where the constraint  $\mathbb{E}(Y|X) = X$  means:  $\mathbb{E}(Y|X = x) = x$  for  $\mu$ -almost every  $x \in \mathbb{R}$ .

We need to define  $W^{\mathbb{R}^2}$ , the Kantorovich metric on  $\mathbb{R}^2$  in order to compare transport plans. It is defined identically to the 1-dimensional version in (1) and (2), except that  $|\cdot|$  is replaced with a norm  $\|\cdot\|$  of  $\mathbb{R}^2$ . Indeed the choice of a norm is required in the definition of the 1-Lipschitzian functions in (1) and more directly in (2). For  $d = 1, 2$ , we denote by  $\mathcal{T}_1(\mathbb{R}^d)$  the topology induced by  $W^{\mathbb{R}^d}$  and  $\mathcal{T}_{\text{cb}}(\mathbb{R}^d)$  the usual weak topology. The letters “cb” stand for continuous bounded functions because they define the weak topology while the former topology is induced by the continuous functions growing at most linearly at infinity.

#### 1.1. Seven partial orders on $\mathcal{M}$

We introduce seven partial orders on  $\mathcal{M}$ , investigate their interdependencies, and explain their meanings in terms of couplings. These definitions will be useful for a synthetic formulation in Section 2, like for instance in Lemma 2.27. The results of this section continue the extension of the convex order started with the extended order in [2] to other cones of functions. They are used in Section 2 but may also be interesting for their own sake. Even if the results

like Theorem 1.7 and Corollary 1.8 may sound classical and the proofs are easy, they are to our knowledge the first appearance in the mathematical literature.

**Definition 1.1.** The letter  $E$  is a variable for a set of real functions growing at most linearly at  $-\infty$  and  $+\infty$ . We introduce the set of non-negative functions  $E_+$ , the set of non-increasing functions  $E_{\text{sto}}$  and the set of convex functions  $E_C$ , all three are restricted to functions with the growing constraint. For  $\mu, \nu \in \mathcal{M}$  we introduce the property  $P(E)$ .

$$P(E) : \forall \phi \in E, \int \phi \, d\mu \leq \int \phi \, d\nu.$$

For  $\mu, \nu \in \mathcal{M}$ ,

- if  $P(E_+)$  holds, we write  $\mu \preceq_+ \nu$  (usual order),
- if  $P(E_{\text{sto}})$  holds, we write  $\mu \preceq_{\text{sto}} \nu$  (stochastic order or first order stochastic dominance),
- if  $P(E_C)$  holds, we write  $\mu \preceq_C \nu$  (convex order, Choquet order or second order stochastic dominance),
- if  $P(E_C \cap E_+)$  holds, we write  $\mu \preceq_{C,+} \nu$ ,
- if  $P(E_+ \cap E_{\text{sto}})$  holds, we write  $\mu \preceq_{+, \text{sto}} \nu$ ,
- if  $P(E_C \cap E_{\text{sto}})$  holds, we write  $\mu \preceq_{C, \text{sto}} \nu$ ,
- if  $P(E_C \cap E_+ \cap E_{\text{sto}})$  holds, we write  $\mu \preceq_{C,+, \text{sto}} \nu$ .

**Remark 1.2 (Usual notations).** The usual notation for  $\mu \preceq_+ \nu$  is  $\mu \leq \nu$ . In [11],  $\stackrel{D}{\leq}$  is the notation for the stochastic order  $\preceq_{\text{sto}}$ . In [8], the author simply denotes  $\preceq_{C, \text{sto}}$  by  $<$ . In [2], Beiglböck and the author introduced the extended order  $\preceq_E$ . The latter is the same as  $\preceq_{C,+}$  in this paper.

### 1.2. Complements on the stochastic order

Recall that the Lebesgue measure is denoted by  $\lambda$ . For a positive measure  $\nu$ , we note  $F_\nu$ , the cumulative distribution function and  $G_\nu$ , the quantile function. Recall that  $G_\nu(t) = \inf_{x \in \mathbb{R}} \{F_\nu(x) \geq t\}$ . This function can be seen as a general inverse of  $F_\nu$ . It is left-continuous and defined on  $[0, \nu(\mathbb{R})]$ . Recall also  $\nu = (G_\nu)_\# \lambda|_{[0, \nu(\mathbb{R})]}$ , which will be used extensively in this paper.

The following standard proposition can for instance be found in [11, Theorem 3.1]. See also the introduction of Section 1.3. The proof makes use of the quantile functions.

**Proposition 1.3.** For  $\mu, \nu \in \mathcal{P}$ , the relation  $\mu \preceq_{\text{sto}} \nu$  holds, if and only if there exists a pair of random variables  $(X, Y)$  on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  with marginals  $\mu$  and  $\nu$ , such that  $X \leq Y$ ,  $\mathbb{P}$ -almost surely.

We can actually choose  $\mathbb{P} = \lambda|_{[0,1]}$ ,  $X = G_\mu$  and  $Y = G_\nu$ . Furthermore, note that with this representation the pair  $(X, Y)$  gives the minimal value in (2). Indeed the bound  $|\mathbb{E}(Y) - \mathbb{E}(X)| \leq \mathbb{E}(|Y - X|)$  is always satisfied but if  $X \leq Y$  we also have  $\mathbb{E}(|Y - X|) = \mathbb{E}(Y) - \mathbb{E}(X)$ . Actually, we have more generally.

**Lemma 1.4.** Let  $\mu, \nu$  be in  $\mathcal{P}$ . Then the coupling  $(G_\mu, G_\nu)$  defined on  $\Omega = ([0, 1], \mathcal{B}, \lambda)$  is optimal in the definition (2) of  $W(\mu, \nu)$ . More generally if  $\mu, \nu$  have mass  $m \neq 1$  we also have

$$W(\mu, \nu) = \int |G_\nu - G_\mu| \, d\lambda_{[0,m]} = \|G_\nu - G_\mu\|_1.$$

Moreover if  $\mu \preceq_{\text{sto}} \nu$ ,

$$W(\mu, \nu) = \int_0^m (G_\nu - G_\mu) \, d\lambda = m \left( \frac{1}{m} \int x \, d\mu - \frac{1}{m} \int x \, d\nu \right).$$

Let us define the *rightmost* and *leftmost measure of mass  $\alpha$  smaller than  $\nu$* . Denoting the mass of  $\nu$  by  $m$  and assuming  $\alpha \leq m$ , we consider the set  $S = \{\mu \in \mathcal{M} \mid \mu(\mathbb{R}) = \alpha \text{ and } \mu \preceq_+ \nu\}$ . Let us prove that for any  $\mu \in S$ , we have

$\mu \preceq_{\text{sto}} \nu_\alpha$  where  $\nu_\alpha$  denotes  $(G_\nu)_{\#\lambda}|_{[m-\alpha, m]}$ . Let  $\mu$  be such a measure and  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  a non-decreasing function, integrable for the elements of  $\mathcal{M}$ . The measure  $\nu_\alpha$  admits a density with respect to  $\nu$  that is 0 on  $]-\infty, G_\nu(m-\alpha)[$  and 1 on  $]G_\nu(m-\alpha), +\infty[$ . The density of  $\mu$  with respect to  $\nu$  is a function with values in  $[0, 1]$ . Hence  $\mu - (\mu \wedge \nu_\alpha)$  is concentrated on  $]-\infty, G_\nu(m-\alpha)[$  and  $\nu_\alpha - (\mu \wedge \nu_\alpha)$  is concentrated on  $]G_\nu(m-\alpha), \infty[$ . As  $G_\nu$  is a non-decreasing function and the mass of both  $\mu$  and  $\nu_\alpha$  is  $\alpha$ , we obtain

$$\int \varphi d(\mu - \mu \wedge \nu_\alpha) \leq [\alpha - (\mu \wedge \nu_\alpha)(\mathbb{R})] \times \varphi(G_\nu(m-\alpha)) \leq \int \varphi d(\nu_\alpha - \mu \wedge \nu_\alpha).$$

Adding  $\int \varphi d(\mu \wedge \nu_\alpha)$  we obtain  $\int \varphi d\mu \leq \int \varphi d\nu$ . The measure  $\nu_\alpha$  is the rightmost measure of mass  $\alpha$  smaller than  $\nu$ . Symmetrically  $(G_\nu)_{\#\lambda}|_{[0, \alpha]}$  is the leftmost measure.

### 1.3. Complements on the convex order

In [10, Theorem 8], Strassen established a statement on the marginals of  $k$ -dimensional martingales indexed on  $\mathbb{N}$ . For our purposes, we restrict the statement to 1-dimensional martingales with one time-step. This result is related to the convex order  $\preceq_C$  in the same way as Proposition 1.3 is associated with  $\preceq_{\text{sto}}$ . Actually, in particular for more general ordered spaces than  $\mathbb{R}$ , Proposition 1.3 is widely referred to as Strassen's theorem on stochastic dominance. The theorem is attributed to Strassen because of [10]. However, the statement of this result in the paper by Strassen is very elusive. It corresponds to two lines on page 438 after the proof of Theorem 11. See a paper by Lindvall [9], where a proof relying on Theorem 7 by Strassen is restituted with all the details. Therefore, we prefer to reserve the name Strassen's theorem for the domination in convex order and we later call similar results, like Proposition 1.3, Strassen-type theorems.

**Proposition 1.5 (Theorem of Strassen 1).** *For  $\mu, \nu \in \mathcal{P}$ , the relation  $\mu \preceq_C \nu$  holds if and only if there exists a pair of random variables  $(X, Y)$  on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  with marginals  $\mu$  and  $\nu$ , such that  $\mathbb{E}(Y|X) = X$ ,  $\mathbb{P}$ -almost surely.*

In the same article [10, Theorem 9] Strassen states a result on submartingales indexed on two times.

**Proposition 1.6 (Theorem of Strassen 2).** *For  $\mu, \nu \in \mathcal{P}$ , the relation  $\mu \preceq_{C, \text{sto}} \nu$  holds if and only if there exists a pair of random variables  $(X, Z)$  on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  with marginals  $\mu$  and  $\nu$ , such that  $\mathbb{E}(Z|X) \geq X$ ,  $\mathbb{P}$ -almost surely.*

Note that if we introduce  $Y = \mathbb{E}(Z|X)$ , one has  $\mu \preceq_{\text{sto}} \text{Law}(Y)$  and  $\text{Law}(Y) \preceq_C \nu$ . This kind of decomposition will be investigated in the next section.

### 1.4. Strassen-type theorems

Before we state Theorem 1.7, let us clarify a point of notation. One may permute the subscripts of  $\preceq$  without changing the meaning of the partial orders. For instance  $\preceq_{+, \text{sto}, C}$  does not appear in Definition 1.1 but it denotes the same order as  $\preceq_{C, +, \text{sto}}$ . More than one notation for the same object seems useless but the arrangement of the indices makes sense in the following theorem.

**Theorem 1.7 (Chain of relations).** *All the relations of Definition 1.1 are antisymmetric and transitive, making them partial orders.*

*Moreover, for any sequence  $(\mu_i)_{i=0, \dots, n}$  (with  $n = 2$  or  $3$ ) satisfying the relations  $\mu_{i-1} \preceq_{r_i} \mu_i$  for  $i = 1, \dots, n$  one has  $\mu_0 \preceq_{r_1, \dots, r_n} \mu_n$ .*

*Conversely if  $\mu_0 \preceq_{r_1, \dots, r_n} \mu_n$  one can find a sequence  $(\mu_i)_{i=0, \dots, n}$  such that  $\mu_{i-1} \preceq_{r_i} \mu_i$  for every  $i \geq 1$ .*

**Proof.** 1. The transitivity is obvious. For the antisymmetry, it is enough to prove that  $\preceq_{C, +, \text{sto}}$  is antisymmetric. Let  $\mu$  and  $\nu$  satisfy  $\mu \preceq_{C, +, \text{sto}} \nu$  and  $\nu \preceq_{C, +, \text{sto}} \mu$ . Hence integrating with respect to  $\mu$  or  $\nu$  provides the same value for

any function that can be written in all three forms – (i) the difference of two non-negative functions, (ii) the difference of two non-decreasing functions, (iii) the difference of two convex functions. All the three spaces are restricted to functions growing at most linearly at  $\pm\infty$ . Continuous piecewise affine functions with finitely many pieces satisfy the three conditions. Thus  $\mu = \nu$ .

2. The first implication is obvious, the converse statement is not. We have to prove it for twelve different partial orders. For  $\preceq_{C,+}$  (see Remark 1.2) we simply quote [2, Proposition 4.4]. From this, we can easily deduce the statement for  $\preceq_{+,C}$ . We consider  $\mu_0 \preceq_{+,C} \mu_1$ . As the order is the same as  $\preceq_{C,+}$ , we can find  $\mu_1$  with  $\mu_0 \preceq_C \mu_1$  and  $\mu_1 \preceq_+ \mu_2$ . We set  $\mu'_1 = \mu_0 + (\mu_2 - \mu_1)$ . As  $\mu_2 - \mu_1$  is a positive measure one has  $\mu_0 \preceq_+ \mu'_1$ . Let  $\varphi$  be a convex function. Therefore

$$\begin{aligned} \int \varphi d\mu'_1 &= \int \varphi d\mu_0 + \int \varphi d(\mu_2 - \mu_1) \\ &\leq \int \varphi d\mu_1 + \int \varphi d(\mu_2 - \mu_1) \leq \int \varphi d\mu_2, \end{aligned}$$

which means  $\mu'_1 \preceq_C \mu_2$ . The last argument can be used for stating the decomposition of  $\preceq_{+,C,\text{sto}}$  and  $\preceq_{C,+, \text{sto}}$  provided we can prove it for  $\preceq_{C,\text{sto},+}$ . The place of the index “+” does not matter. Similarly the decomposition of  $\preceq_{+, \text{sto},C}$  and  $\preceq_{\text{sto},+,C}$  will be a corollary of the property for  $\preceq_{\text{sto},C,+}$ . In the same way  $\preceq_{+, \text{sto}}$  reduces to the study of  $\preceq_{+, \text{sto}}$ .

3. We prove here the two wanted decompositions of  $\mu \preceq_{C,\text{sto}} \nu$ . For probability measures, the Strassen theorem (Proposition 1.6) states that there exists  $(X, Z)$  with  $\text{Law}(X) = \mu$ ,  $\text{Law}(Z) = \nu$  and  $Y := \mathbb{E}(Z|X) \geq X$ . For  $\mu_1$  defined as the law of  $Y$  and  $\mu'_1$  as the law of  $Y' := Z - (Y - X)$  we have  $\mu = \mu_0 \preceq_{\text{sto}} \mu_1 \preceq_C \mu_2 = \nu$  and  $\mu = \mu_0 \preceq_C \mu'_1 \preceq_{\text{sto}} \mu_2 = \nu$ . If  $\mu, \nu$  are not probability measures, they must have the same mass. Indeed, every constant function is element of  $E_C \cap E_{\text{sto}}$ . One can easily obtain the statement by normalising the measures.

4. We are left with  $\preceq_{\text{sto},+}$ ,  $\preceq_{\text{sto},C,+}$  and  $\preceq_{C,\text{sto},+}$ . Having in mind the possibility to transpose “C” and “sto” proved in the last paragraph, it is sufficient to consider  $\mu \preceq_{\text{sto},+} \nu$  and  $\mu \preceq_{\text{sto},C,+} \nu$ . For that purpose we consider  $\nu' = (G_\nu) \# \lambda|_{[\nu(\mathbb{R}) - \mu(\mathbb{R}), \nu(\mathbb{R})]}$ . Recall that it is the rightmost measure of mass  $\mu(\mathbb{R})$  smaller than  $\nu$  introduced in Section 1.2. Of course  $\nu' \preceq_+ \nu$ . We now prove  $\mu \preceq_{\text{sto}} \nu'$  and  $\mu \preceq_{\text{sto},C} \nu'$  respectively. Let  $\varphi \in E$  with  $E = E_{\text{sto}}$  or  $E = E_{\text{sto}} \cap E_C$  respectively. Because of the dominated convergence theorem, we can assume without loss of generality that  $\varphi$  is bounded from below. We denote  $G_\nu(\nu(\mathbb{R}) - \mu(\mathbb{R}))$  by  $x \in [-\infty, +\infty[$  so that  $\varphi - \varphi(x)$  is non-negative on  $]x, +\infty[$ . For simplicity, one considers that  $\mu$  is a probability measure. By applying  $\mu \preceq_{\text{sto},+} \nu$  or  $\mu \preceq_{\text{sto},C,+} \nu$  for  $(\varphi - \varphi(x))\chi_{]x, +\infty[}$  respectively, one obtains

$$\begin{aligned} \int \varphi d\mu &= \varphi(x) + \int [\varphi - \varphi(x)]\chi_{]x, +\infty[} d\mu \\ &\leq \varphi(x) + \int [\varphi - \varphi(x)]\chi_{]x, +\infty[} d\nu = \int \varphi d\nu'. \end{aligned}$$

Hence  $\mu \preceq_{\text{sto}} \nu'$  and  $\mu \preceq_{\text{sto},C} \nu'$  respectively. For the latter we recall point 3 so that we have  $\mu = \mu_0 \preceq_{\text{sto}} \mu_1 \preceq_C \mu_2 = \nu'$  and  $\mu_2 \preceq_+ \mu_3 = \nu$  for some intermediate measure  $\mu_1$ . □

Theorem 1.7 opens the door for a translation of all the partial orders of Definition 1.1 in terms of couplings. For this purpose we use what is known on  $\preceq_{\text{sto}}$  and  $\preceq_C$  (Proposition 1.3 and Proposition 1.5) together with the following characterisation: if  $\nu \in \mathcal{P}$  then  $\mu \preceq_+ \nu$  if and only if there exists a random variable  $Y$  defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , with  $\text{Law}(Y) = \nu$  and an event  $A$  such that  $\mu(\mathbb{R}) = \mathbb{P}(A)$  and  $\text{Law}(Y|A) = \mu(\mathbb{R})^{-1}\mu$ . The statement also requires the composition of joint laws, called gluing lemma in [13]. As an example let us reprove the converse statement of Proposition 1.6. We start with  $\mu_0, \mu_2 \in \mathcal{P}$  satisfying  $\mu_0 \preceq_{C,\text{sto}} \mu_2$ . With Theorem 1.7, we find  $\mu_1$  satisfying  $\mu_0 \preceq_C \mu_1$  and  $\mu_1 \preceq_{\text{sto}} \mu_2$ . Hence on some probability space  $\Omega_X$  we have a coupling  $(X_0, X_1)$  of  $\mu_0$  and  $\mu_1$  that satisfies  $\mathbb{E}(X_1|X_0) = X_0$  and on some probability space  $\Omega_Y$  we have a coupling  $(Y_1, Y_2)$  of  $\mu_1$  and  $\mu_2$  that satisfies  $Y_1 \leq Y_2$ . Therefore by using the Markov composition, or the gluing lemma [13, Chapter 1], there exists some probability space  $\Omega_Z$  and  $(Z_0, Z_1, Z_2)$  such that  $\text{Law}(Z_0, Z_1) = \text{Law}(X_0, X_1)$  and  $\text{Law}(Z_1, Z_2) = \text{Law}(Y_1, Y_2)$ . It follows  $\mathbb{E}(Z_2|Z_0) \geq \mathbb{E}(Z_1|Z_0) = Z_0$ .

We give another illustration on how to apply Theorem 1.7 in the case of an order made of three subscripts.

**Corollary 1.8.** *Let  $\mu, \nu$  be elements of  $\mathcal{M}$ . Then the relation  $\mu \preceq_{C,+,\text{sto}} \nu$  holds if and only if there exists a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  with a measurable set  $A$  and two random variables  $(X, T)$  satisfying*

$$\mathbb{P}_A\text{-almost surely } X \leq \mathbb{E}(T|X, A),$$

where  $\mathbb{P}_A(\cdot)$  is the conditional probability  $\mathbb{P}(\cdot \cap A)/\mathbb{P}(A)$  and

- (1)  $\text{Law}(X|A) = \mu(\mathbb{R})^{-1}\mu,$
- (2)  $\text{Law}(T) = \nu(\mathbb{R})^{-1}\nu,$
- (3)  $\mathbb{P}(A) = \mu(\mathbb{R})\nu(\mathbb{R})^{-1}.$

**Proof.** 1. According to Theorem 1.7, setting  $\mu_0 = \mu$  and  $\mu_3 = \nu$ , we can find  $\mu_1, \mu_2 \in \mathcal{M}$  with  $\mu_0 \preceq_C \mu_1 \preceq_+ \mu_2 \preceq_{\text{sto}} \mu_3$ . We first assume  $\mu_3 \in \mathcal{P}$  for simplicity. We apply Proposition 1.3 and Proposition 1.5 to the pairs  $(\mu_2, \mu_3)$  and  $(\mu_0, \mu_1)$ . According to the usual composition rules of the probability theory, we can find a pair  $(Z, T)$  for  $(\mu_2, \mu_3)$  and  $(X, Y)$  for  $(\mu_0(\mathbb{R})^{-1}\mu_0, \mu_1(\mathbb{R})^{-1}\mu_1)$  satisfying the relations explained in these propositions. Moreover usual properties of the probability theory allow us to couple these random variables in a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and its restriction  $(A, \mathcal{F}_A, \mathbb{P}_A)$  where  $A \subseteq \Omega$  is a Borel set adapted to the relation  $\mu_1 \preceq_+ \mu_2$ . It satisfies  $\mathbb{P}(A)\mu_2(\mathbb{R}) = \mu_1(\mathbb{R})$  and we have

- $\text{Law}(T) = \mu_3,$
- $\text{Law}(Z) = \mu_2,$
- $\text{Law}(Y) = \mathbb{P}(A)^{-1}\mu_1,$
- $\text{Law}(X) = \mathbb{P}(A)^{-1}\mu_0$

and

- $Z \leq T,$
- $Y = Z, \mathbb{P}_A\text{-almost surely},$
- $X = \mathbb{E}_A(Y|X), \mathbb{P}_A\text{-almost surely}.$

The last line also writes  $X = \mathbb{E}(Z|X, A), \mathbb{P}_A\text{-almost surely}$ . Thus  $X \leq \mathbb{E}(T|X, A), \mathbb{P}_A\text{-almost surely}$ .

2. We prove the converse statement. We assume that  $\mathbb{P}_A\text{-almost surely } X \leq \mathbb{E}(T|X, A)$  is satisfied and consider  $\varphi \in E_C \cap E_+ \cap E_{\text{sto}}$ . We have  $\mathbb{E}_A(\varphi(X)) \leq \mathbb{E}_A(\varphi(\mathbb{E}(T|X, A)))$  because  $\varphi$  is non-decreasing. This is smaller than  $\mathbb{E}_A(\varphi(T))$  because  $\varphi$  is convex. Finally this is smaller than  $\mathbb{P}(A)\mathbb{E}(\varphi(T))$  because  $\varphi$  is non negative. We conclude with (1)–(3) that  $\int \varphi d\mu \leq \int \varphi d\nu$ .

3. The statement is established if  $\nu = \mu_3$  is a probability measure. Using the usual normalisation of positive measures to probability measures, we get the other cases. □

## 2. Lipschitz continuity of the curtain coupling with respect to its marginals

In this section we recall the properties of the martingale curtain coupling  $\pi_{1c} = \text{Curt}(\mu, \nu)$  between two measures  $\mu \preceq_C \nu$ . We prove that it is a continuous map by using the property of monotonicity satisfied by curtain couplings. We establish a Lipschitz estimate for the shadow projection  $(\mu, \nu) \mapsto S^\nu(\mu)$  and deduce that  $\text{Curt} : (\mu, \nu) \in \mathcal{P} \times \mathcal{P} \longrightarrow \Pi_M$  is Lipschitzian when  $\Pi_M$  is considered with the ad hoc (semi)metric  $Z$ . We also prove that such an estimate does not hold in  $(\Pi_M, W^{\mathbb{R}^2})$ . An important mathematical object introduced in this section is the reduced support that we denote  $\text{spt}^* \pi$ . This set of full mass contributes to a better understanding of the property of monotonicity.

### 2.1. Definitions of the shadows and the curtain coupling

In [2, Lemma 4.6] the following important theorem-definition is proven.

**Definition 2.1 (Definition of the shadow).** *If  $\mu \preceq_{C,+} \nu$ , there exists a unique measure  $\eta$  such that*

- $\mu \preceq_C \eta,$
- $\eta \preceq_+ \nu,$

- if  $\eta'$  satisfies the two first conditions (i.e  $\mu \preceq_C \eta' \preceq_+ \nu$ ), one has  $\eta \preceq_C \eta'$ .

This measure  $\eta$  is called the shadow of  $\mu$  in  $\nu$  and we denote it by  $S^\nu(\mu)$ .

The shadows are sometimes difficult to determine. An important fact is that they have the smallest variance among the set of measures  $\eta'$ . Indeed,  $\eta \preceq_C \eta'$  implies  $\int x \, d\eta = \int x \, d\eta'$  and  $\int x^2 \, d\eta \leq \int x^2 \, d\eta'$  with equality if and only if  $\eta = \eta'$  or  $\int x^2 \, d\eta = +\infty$ .

**Example 2.2 (Shadow of an atom, Example 4.7 in [2]).** Let  $\delta$  be an atom of mass  $\alpha$  at a point  $x$ . Assume that  $\delta \preceq_{C,+} \nu$ . Then  $S^\nu(\delta)$  is the restriction of  $\nu$  between two quantiles, more precisely it is  $\nu' = (G_\nu)_{\#} \lambda_{s',s'[\}$  where  $s' - s = \alpha$  and the barycenter of  $\nu'$  is  $x$ .

The next lemma describes the tail of the shadows.

**Lemma 2.3.** Let  $\mu, \nu \in \mathcal{M}$  satisfy  $\mu \preceq_{C,+} \nu$ . Assume that  $y = \sup[\text{spt } \mu]$  is finite. Then the restriction of  $(S^\nu(\mu) - \mu)_+$  to  $[y, +\infty[$  is the stochastically leftmost measure  $\theta$  among the measures of the same mass satisfying  $\theta \preceq_+ (\nu - \mu)_+|_{[y, +\infty[}$ .

The corresponding statement holds in the case  $\inf[\text{spt } \mu] > -\infty$ .

Before we write the proof, let us make clear that except if  $y$  is an atom of both  $\mu$  and  $\nu$ , the measure  $(S^\nu(\mu) - \mu)_+|_{[y, +\infty[}$  is simply  $S^\nu(\mu)|_{[y, +\infty[}$  and  $(\nu - \mu)_+|_{[y, +\infty[}$  is simply  $\nu|_{[y, +\infty[}$ .

**Proof.** Using Strassen’s theorem (Proposition 1.5), let  $\pi$  be a martingale transport plan with marginals  $\mu$  and  $S^\nu(\mu)$ . Let  $(\pi_x)_{x \in \mathbb{R}}$  be a disintegration where the measures  $\pi_x$  are probability measures. Each  $\pi_x$  can again be disintegrated in a family of probability measures concentrated on two points and with barycenter  $x$ . Observe now that for  $a < x < b$  and  $b' \in ]x, b]$ , one can compare  $\frac{b-x}{b-a} \delta_a + \frac{x-a}{b-a} \delta_b$  with  $\frac{b'-x}{b'-a} \delta_a + \frac{x-a}{b'-a} \delta_{b'}$  in the following way:

- both measures have mass 1 and barycenter  $x$ ,
- $\frac{b'-x}{b'-a} \delta_a \preceq_+ \frac{b-x}{b-a} \delta_a$  (inequality for the mass in  $a$ ),
- $\frac{b'-x}{b'-a} \delta_a + \frac{x-a}{b'-a} \delta_{b'} \preceq_C \frac{b-x}{b-a} \delta_a + \frac{x-a}{b-a} \delta_b$ .

Remind that  $S^\nu(\mu) = \int [\frac{b-x}{b-a} \delta_a + \frac{x-a}{b-a} \delta_b] \, d\zeta_0(x, a, b)$  where  $\zeta_0$  is a positive measure with first marginal  $\mu$  that is concentrated on  $\{(x, a, b) \in \mathbb{R}^3, a < x < b \text{ or } a = x = b\}$ . For  $a = x = b$ , we adopt the convention  $\frac{b-x}{b-a} \delta_a + \frac{x-a}{b-a} \delta_b = \delta_x$ . The measure  $(S^\nu(\mu) - \mu)_+|_{[y, +\infty[}$  of the statement can be written  $\theta = \int [\frac{x-a}{b-a} \delta_b] \, d\zeta(x, a, b)$  where  $\zeta \preceq_+ \zeta_0$  and  $\zeta$  is concentrated on  $\{(x, a, b) \in \mathbb{R}^3, x < b\}$ . Let  $\theta'$  satisfy  $\theta' \preceq_+ (\nu - \mu)_+|_{[y, +\infty[}$  and  $\theta' \preceq_{\text{sto}} \theta$ . Hence one can consider a measure  $\bar{\zeta}$  concentrated on  $\{(x, a, b, b') \in \mathbb{R}^4, x \leq b' < b\}$  such that  $\theta' = \int [\frac{x-a}{b'-a} \delta_{b'}] \, d\bar{\zeta}(x, a, b, b')$  and the projection of  $\bar{\zeta}$  on the three first coordinates is  $\zeta$ . We denote by  $\zeta'$  the measure  $\frac{b'-a}{b-a} \bar{\zeta}$  and with a slight abuse of notation we denote also by  $\zeta'$  its projection on the first three coordinates. We set

$$\eta = \int \left[ \frac{b'-x}{b'-a} \delta_a + \frac{x-a}{b'-a} \delta_{b'} \right] \, d\zeta'(x, a, b, b') + \int \left[ \frac{b-x}{b-a} \delta_a + \frac{x-a}{b-a} \delta_b \right] \, d(\zeta_0 - \zeta')(x, a, b).$$

Recall that

$$S^\nu(\mu) = \int \left[ \frac{b-x}{b-a} \delta_a + \frac{x-a}{b-a} \delta_b \right] \, d\zeta'(x, a, b, b') + \int \left[ \frac{b-x}{b-a} \delta_a + \frac{x-a}{b-a} \delta_b \right] \, d(\zeta_0 - \zeta')(x, a, b).$$

Therefore according to the three remarks above, one has

- $\mu \preceq_C \eta$ ,



- $\eta \leq_+ \nu$ ,
- $\eta \leq_C S^\nu(\mu)$ .

The second relation relies on

$$\int \left[ \frac{b' - x}{b' - a} \delta_a + \frac{x - a}{b' - a} \delta_{b'} \right] d\zeta'(x, a, b, b') = \int \left[ \frac{b' - x}{b - a} \delta_a \right] d\zeta' + \theta'.$$

The last relation is in fact an equality. Indeed, the domination  $\eta \geq_C S^\nu(\mu)$  is a consequence of the two first relations and the definition of the shadow. Moreover  $\leq_C$  is antisymmetric so that  $\eta = S^\nu(\mu)$ . Hence  $\zeta'$ -almost surely we have  $b = b'$ , which means  $\theta' = \theta$ . We have proven that the restriction of  $(S^\nu(\mu) - \mu)_+$  to  $]y, +\infty[$  is the stochastically leftmost measure smaller than  $(\nu - \mu)_+|_{]y, +\infty[}$ .  $\square$

The following result is one of the most important on the structure of shadows. It is Theorem 4.8 of [2].

**Proposition 2.4 (Structure of shadows).** *Let  $\gamma_1, \gamma_2$  and  $\nu$  be elements of  $\mathcal{M}$  and assume that  $\mu = \gamma_1 + \gamma_2 \leq_{C,+} \nu$ . Then we have  $\gamma_2 \leq_{C,+} \nu - S^\nu(\gamma_1)$  and*

$$S^\nu(\gamma_1 + \gamma_2) = S^\nu(\gamma_1) + S^{\nu - S^\nu(\gamma_1)}(\gamma_2).$$

**Example 2.5 (Shadow of a finite sum of atoms).** *Let  $\mu$  be the measure  $\sum_{i=1}^n \alpha_i \delta_{x_i}$  and  $\nu = G_{\#\lambda}]0, m]$  such that  $\mu \leq_{C,+} \nu$ . We can apply Proposition 2.4 to this sum as well as Example 2.2 on the shadow of one atom. We obtain recursively the following description. There exists an increasing sequence of sets  $J_1 \subseteq \dots \subseteq J_n \subseteq ]0, m]$  satisfying that  $J_k$  has measure  $\sum_{i=1}^k \alpha_i$  and  $J_k \setminus J_{k-1}$  is a pseudo-interval of  $]0, m] \setminus J_{k-1}$ , that is  $J_k \setminus J_{k-1} = ]s, t] \setminus J_{k-1}$  for some  $0 \leq s, t \leq m$ . These pseudo-intervals satisfy  $S^\nu(\sum_{i=1}^k \alpha_i \delta_{x_i}) = G_{\#\lambda_{J_k}}$  for every  $k \leq n$ .*

*Conversely any increasing sequence  $(J_i)_{i=1, \dots, n}$  such that  $J_k \setminus J_{k-1}$  is a pseudo-interval of  $]0, m] \setminus J_{k-1}$  is associated with a family of atoms  $\alpha_i \delta_{x_i}$  with  $\alpha_i = \lambda(J_i) - \lambda(J_{i-1})$  and  $x_i$  is the barycenter of  $G_{\#\lambda_{J_i \setminus J_{i-1}}}$  such that  $G_{\#\lambda_{J_k}}$  is the shadow of  $\sum_{i=1}^k \alpha_i \delta_{x_i}$  in  $\nu$ .*

With the shadow projections, we can introduce the left-curtain coupling. For atomic measures it is related to Example 2.5 when we assume that  $(x_i)_i$  is an increasing sequence.

**Definition 2.6 (Left-curtain coupling, Theorem 4.18 in [2]).** *Let  $\mu, \nu \in \mathcal{M}$  satisfy  $\mu \leq_C \nu$ . There exists a unique measure  $\pi \in \Pi_M(\mu, \nu)$  such that for any  $x \in \mathbb{R}$  the measure  $\pi|_{]-\infty, x] \times \mathbb{R}}$  has first marginal  $\mu|_{]-\infty, x]}$  and second marginal  $S^\nu(\mu|_{]-\infty, x]})$ . We denote it by  $\pi|_C$  and call it left-curtain coupling.*

One of the main theorems of [2] is the equivalence of three properties of couplings: left-curtain, left-monotone and optimal. Let us define left-monotone couplings.

**Definition 2.7 (Left-monotone coupling).** *Let  $\pi$  be an element of  $\Pi_M(\mu, \nu)$ . The coupling  $\pi$  is left-monotone if there exists a Borel set  $\Gamma$  with*

- $\pi(\Gamma) = 1$ ,
- for every  $(x, y^-), (x, y^+)$  and  $(x', y')$  elements of  $\Gamma$  satisfying  $x < x'$  and  $y^- < y^+$ , the real  $y'$  is not an element of  $]y^-, y^+[$ .

We can now state the result.

**Proposition 2.8 (Theorem 1.9 in [2]).** *Let  $\pi \in \Pi_M(\mu, \nu)$ . We introduce  $c : (x, y) \in \mathbb{R}^2 \rightarrow [1 + \tanh(-x)]\sqrt{y^2 + 1}$ . Then the properties are equivalent.*

- *Left-curtain: the transport plan  $\pi$  is the left-curtain coupling,*
- *Left-monotone: the transport plan  $\pi$  is left-monotone,*
- *Optimal: for any  $\tilde{\pi} \in \Pi_M(\mu, \nu)$ , if  $\tilde{\pi} \neq \pi$ , then  $\int c d\pi < \int c d\tilde{\pi}$ .*

**Remark 2.9.** See Example 2.11 about the fact that the left-monotonicity may not be satisfied for  $\Gamma = \text{spt } \pi$  even if it is realised for another  $\Gamma$ .

**Remark 2.10.** Actually Theorem 1.9 in [2] is written for another kind of cost  $c$ . However replacing Theorem 6.1 by Theorem 6.3, both of this paper, leads to this version. Actually the latter theorem states that if  $c$  is defined as  $(x, y) \mapsto \varphi(x)\psi(y)$  where  $\varphi$  is positive and decreasing,  $\psi$  is positive and strictly convex the implication “optimal  $\Rightarrow$  left-curtain” still holds provided  $\min_{\tilde{\pi} \in \Pi_M(\mu, \nu)} \int c \, d\tilde{\pi}$  is finite. In Proposition 2.8 this condition is satisfied without more assumptions because  $\mu, \nu$  have finite first moments and the given  $c$  grows at most linearly at  $\pm\infty$ .

In [5], Henry-Labordère and Touzi have proved that functions  $c$  such that the partial derivative  $\partial_{yxx} c$  is identically negative also lead to the left-curtain coupling if  $\min_{\tilde{\pi} \in \Pi_M(\mu, \nu)} \int c \, d\tilde{\pi}$  is finite. This contains, in the case of smooth functions  $c$ , both the kind of costs in [2, Theorem 6.1] and [2, Theorem 6.3].

### 2.2. Qualitative continuity of the curtain coupling map

In this paragraph we show that  $\text{Curt} : (\mu, \nu) \mapsto \pi_{1c}$  is continuous on  $\mathcal{P}^2$ . For the proof we are using the second characterisation of left-curtain couplings: according to Proposition 2.8 they are exactly the *left-monotone* couplings defined in Definition 2.7 with a set  $\Gamma \subseteq \mathbb{R}^2$  of full mass. Recall the classical definition of the support of a measure: the support  $\text{spt}(\pi)$  is the smallest closed set  $C \subseteq \mathbb{R}^2$  such that  $\pi(\mathbb{R}^2 \setminus C) = 0$ . More generally if  $\pi(\mathbb{R}^2 \setminus C' = 0)$  for some measurable set  $C'$ , we say that  $\pi$  is concentrated on  $C'$ . Example 2.11 illustrates that for a left-monotone  $\pi$  the set  $\Gamma = \text{spt}(\pi)$  may not fulfill the desired properties in Definition 2.7, which contrasts with the theory of the classical transport problem. The measure  $\pi$  is concentrated on the modified support  $\text{spt}^*(\pi)$  that we define after Example 2.11. We call it “support” even if it is not necessarily a closed set. In fact as  $\text{spt}^*(\pi) \subseteq \text{spt}(\pi)$ , it is a closed set if and only if it is the support of  $\pi$ . It is proved in Proposition 2.14 that unlike  $\text{spt}(\pi)$ , the modified support satisfies the desired properties in Definition 2.7.

**Example 2.11.** Consider  $\mu = (1/2)\lambda_{[-1,1]}$  and  $\nu = (\delta_{-1} + 2\delta_0 + \delta_1)/4$ . For these marginals, considering the transport plan given by the left-curtain coupling, the mass contained in  $[-1, 0]$  is mapped to  $\{-1, 0\}$  while the mass in  $[0, 1]$  is mapped to  $\{0, 1\}$ . Thus  $(0, -1)$ ,  $(0, 1)$  and  $(1, 0)$  are elements of  $\text{spt}(\pi)$ .

This example shows a typical difficulty that may arise on the diagonal set  $\{(x, y) \in \text{spt}(\pi), y = x\}$ , for instance at the *cut points*: those points  $x$  satisfying  $u_\mu(x) = u_\nu(x)$ . According to [2, Lemma 8.5] these cut points provide a decomposition of the martingale transport plans in irreducible components of  $\mathbb{R}$  that depends only on  $\mu$  and  $\nu$ . In particular  $\pi'([-\infty, x]^2 \cup [x, +\infty]^2) = 1$  for every  $\pi' \in \Pi_M(\mu, \nu)$ . In Example 2.11,  $x = 0$  is such a cut point. Proposition 2.14 shows that the problem raised in Example 2.11 does not occur when considering  $\text{spt}^*(\pi)$  instead of  $\text{spt}(\pi)$ .

First, let  $A$  be the set of  $x \in \mathbb{R}$  such that  $\pi([-\infty, x[ \times ]x, \infty]) = 0$ . Note for completeness that the cut points are elements of  $A$ . Second, we denote the subset of  $A$  of points that are isolated in  $A$  on the right by  $A^-$ . Note that  $A^-$  is countable. Finally we set

$$\text{spt}^*(\pi) = (\text{spt}(\pi) \setminus (A^- \times \mathbb{R})) \cup \bigcup_{\mu(x) > 0} \{x\} \times \text{spt } \pi_x,$$

where the kernel  $(\pi_x)_{x \in \mathbb{R}}$  arises from the disintegration with respect to the projection on the first variable. As usual it is only  $\mu$ -almost surely uniquely defined. However for every atom  $x$  of  $\mu$ , the conditional measure  $\pi_x$  is uniquely determined so that the definition of  $\text{spt}^*(\pi)$  does not suffer from any ambiguity. Moreover for  $\mu(x) > 0$  we have always  $\{x\} \times \text{spt}(\pi_x) \subseteq \text{spt}(\pi)$  so that  $\text{spt}^*(\pi) \subseteq \text{spt}(\pi)$ . Using the additivity of  $\pi$  and the fact that  $A^-$  is countable we obtain that  $\pi$  is concentrated on  $\text{spt}^*(\pi)$ .

$$\begin{aligned} \pi(\text{spt}^* \pi) &= [\pi(\text{spt}(\pi) \setminus (A^- \times \mathbb{R}))] + \pi \left[ \left( \bigcup_{\mu(x) > 0} \{x\} \times \text{spt } \pi_x \right) \cap (A^- \times \mathbb{R}) \right] \\ &= [1 - \pi(A^- \times \mathbb{R})] + \sum_{x \in A^-, \mu(x) > 0} \pi(\{x\} \times \text{spt}(\pi_x)) = 1 - \mu(A^-) + \sum_{x \in A^-} \mu(x) = 1. \end{aligned}$$

In Proposition 2.14 and Theorem 2.16 we will use many times Lemma 2.13. This lemma relies itself on the following statement.

**Lemma 2.12.** *Let  $(x, y) \in \text{spt } \pi$  where  $\pi$  is a martingale transport plan and  $G$  a Borel set such that  $\pi(G) = 1$ .*

*If  $x < y$ , for any  $\varepsilon > 0$  there exist  $(x_1, y_1^-), (x_1, y_1^+) \in G$  with  $y_1^- \leq y_1^+$ , such that the point  $(x_1, y_1^+)$  is in the ball of centre  $(x, y)$  and radius  $\varepsilon$  and  $y_1^- < x + \varepsilon$ .*

*If  $x > y$ , the symmetric statement holds as well. There exists  $(x_1, y_1^-), (x_1, y_1^+) \in G$  with  $y_1^- \leq y_1^+$ , such that the point  $(x_1, y_1^-)$  is in the ball of centre  $(x, y)$  and radius  $\varepsilon$  and  $y_1^+ > x - \varepsilon$ .*

**Proof.** It is sufficient to prove the first statement because the second is proved in the same way. We also can assume without loss of generality that  $x < y - \varepsilon$ . We consider the usual disintegration of  $\pi$  with respect to  $\mu = \text{proj}_\#^x \pi$ . Let us denote the vertical cut  $G \cap (\{s\} \times \mathbb{R})$  by  $\{s\} \times G_s$ . For  $\mu$ -almost every  $s$ , we have  $\pi_s(G_s) = 1$  and the expectation of  $\pi_s$  is  $s$ . Moreover as  $x$  is in the support of  $\mu$ , and  $(x, y)$  in the support of  $\pi$ , we have also  $\pi_s ]y - \varepsilon, y + \varepsilon[ > 0$ , for any  $s$  in a subset  $S \subseteq ]x - \varepsilon, x + \varepsilon[$  of positive  $\mu$ -measure. As  $x \notin ]y - \varepsilon, y + \varepsilon[$ , for almost every element  $s \in S$ , we have  $\max(x - \varepsilon, \inf \text{spt } \pi_s) < s < \min(x + \varepsilon, \sup \text{spt } \pi_s)$ . Hence, we can find  $(x_1, y_1^-)$  and  $(x_1, y_1^+)$  in  $G$  with  $\max(|x - x_1|, |y - y_1^+|) \leq \varepsilon$  and  $y_1^- < x + \varepsilon$ .  $\square$

**Lemma 2.13.** *Let  $\pi$  be a martingale transport plan,  $(x, y)$  and  $(x', y')$  elements of  $\text{spt}(\pi)$  and assume  $x < x'$ . If  $x < y' < y$  or  $x > y' > y$ , then  $\pi$  is not left-monotone.*

**Proof.** It is sufficient to prove the first statement for  $x < y' < y$ ; the other case is similar. Let  $(x, y)$  and  $(x', y')$  be elements of  $\text{spt}(\pi)$  with  $x < y' < y$  and  $\Gamma$  have  $\pi$ -measure 1. We apply Lemma 2.12 to  $G = \Gamma$  and  $\varepsilon < \min(|y' - x|, |y' - y|, |x' - x|)$  and find  $x_1, y_1^-$  and  $y_1^+$  with  $(x_1, y_1^\pm) \in \Gamma$ ,  $|x_1 - x| < \varepsilon$ ,  $|y_1^+ - y| < \varepsilon$  and  $y_1^- < x + \varepsilon$ . We are in the forbidden configuration appearing in the definition of  $\Gamma$  because  $x_1 < x'$  and  $y_1^- < y' < y_1^+$ . This is not directly a contradiction because  $(x', y')$  may not be an element of  $\Gamma$ . Nevertheless, as  $(x', y') \in \text{spt}(\pi) \subseteq \bar{\Gamma}$  we can replace it with some element  $(x'_1, y'_1) \in \Gamma$ . As  $\Gamma$  was an arbitrary set,  $\pi$  is not left-monotone.  $\square$

**Proposition 2.14.** *A martingale transport plan  $\pi$  is the left-monotone coupling of  $\Pi_M(\mu, \nu)$  if and only if it satisfies the following condition*

- for every  $(x, y^+), (x, y^-)$  and  $(x', y')$  elements of  $\text{spt}^*(\pi)$ , if  $x < x'$  and  $y^- < y^+$ , we have  $y' \notin ]y^-, y^+[$ .

**Proof.** Let us first prove that  $\pi$  is left-monotone. The set  $\text{spt}^*(\pi)$  fulfills the requirements for  $\Gamma$ . Indeed  $\pi(\text{spt}^*(\pi)) = 1$  and the second condition is assumed in the statement.

Conversely, we assume now that there exists some  $\Gamma$  of mass 1 that satisfies the conditions in Definition 2.7. Without loss of generality, we can assume  $\Gamma \subseteq \text{spt}(\pi)$ : just take  $\Gamma \cap \text{spt}(\pi)$ . By contradiction we consider  $(x, y^+)$  and  $(x, y^-)$  and  $(x', y')$  in  $\text{spt}^*(\pi)$  such that  $x < x'$  and  $y' \in ]y^-, y^+[$ . Note that  $\text{spt}^*(\pi) \subseteq \text{spt}(\pi) \subseteq \bar{\Gamma}$ . In particular each point of  $\text{spt}^*(\pi)$  can be approximated by a point of  $\Gamma$ .

We distinguish two cases.

*Case 1:  $y' \neq x$ .* We can easily conclude applying Lemma 2.13 to  $(x', y')$  and  $(x, y^+)$  or  $(x, y^-)$  depending respectively whether  $x < y'$  or  $x > y'$ .

*Case 2:  $y' = x$ .* We can also assume  $\pi ]-\infty, x[ \times ]x, +\infty[ = 0$  because if not there exists  $(x_1, y_1) \in \text{spt } \pi$  with  $x_1 < x$  and  $x_1 < y' < y_1$ , which permits us to apply Lemma 2.13 to  $(x', y')$  and  $(x_1, y_1)$  and provides a contradiction with the fact that  $\pi$  is left-monotone. Hence we have  $x \in A$ . Remind that  $(x, y^+) \in \text{spt } \pi$  and  $y^+ > y' = x$ . Therefore  $x$  is isolated from the right in  $A$ , so  $x \in A^-$ . According to the definition of  $\text{spt}^*(\pi)$  it implies  $\pi(\{x\} \times \mathbb{R}) = \mu(x) > 0$  and  $y^-, y^+ \in \text{spt}(\pi_x)$ . As  $\mu(x) > 0$  we must have  $\pi_x(\Gamma_x) = 1$  where  $\Gamma_x = \{y \in \mathbb{R}, (x, y) \in \Gamma\}$ . Hence we can find two points  $y_1^\pm \in \Gamma_x$  that are close to  $y^\pm$ . The points  $(x, y_1^-), (x, y_1^+)$  together with some point of  $\Gamma$  close to  $(x', y')$  provide a contradiction.  $\square$

**Corollary 2.15.** *A martingale transport plan  $\pi$  is the left-monotone coupling of  $\Pi_M(\mu, \nu)$  if and only if it satisfies the following condition*

- for every  $(x^-, y^+), (x^+, y^-)$  and  $(x', y')$  elements of  $\text{spt}^*(\pi)$ , if  $x^- \leq x^+ < x'$  and  $y^- < y^+$ , then we have  $y' \notin ]y^-, y^+[$ .

**Proof.** If  $\pi$  satisfies this condition, it also satisfies the sufficient condition in Proposition 2.14 so that it is the left-curtain coupling. For the other inclusion, let us consider  $\pi$  and three points  $(x^-, y^+)$ ,  $(x^+, y^-)$  and  $(x', y')$  in  $\text{spt}^*(\pi)$  as in the statement but such that  $y' \in ]y^-, y^+[$ . We prove that it is not a left-curtain coupling. If  $x^- = x^+$ , we simply use the necessary condition in Proposition 2.14. Hence we assume  $x^- < x^+$ . We can now apply Lemma 2.13 with  $(x', y')$  and  $(x^-, y^+)$  or  $(x^+, y^-)$  depending on whether  $x^- < y'$  or  $x^+ > y'$ .  $\square$

With the last statement we can now implement the strategy of Theorem 5.20 in [13] in order to prove the continuity of the curtain coupling Curt. Actually with Corollary 2.15 it is by now possible to consider triples of points that are typical for the measure  $\pi^{\otimes 3}$  instead of vectors  $(x, y^-, y^+, x', y')$  in  $\mathbb{R}^5$ .

Recall that in paragraph 2.3 we will prove the Lipschitz continuity of Curt for a specific semimetric  $Z$  by using another method.

**Theorem 2.16.** *Consider the mapping  $\text{Curt} : (\mu, \nu) \in \mathcal{D}_{\leq C} \mapsto \pi_{1c}$ , where  $\mathcal{D}_{\leq C} = \{(\mu, \nu) \in \mathcal{P}^2 : \mu \leq_C \nu\}$ . This mapping is continuous from  $\mathcal{D}_{\leq C} \subseteq \mathcal{P}^2$  to  $\Pi_M \subseteq \mathcal{P}(\mathbb{R}^2)$  where the spaces are equipped with the topologies  $\mathcal{T}_{\text{cb}}(\mathbb{R}^2)$  and  $\mathcal{T}_{\text{cb}}(\mathbb{R}^2)$ .*

**Remark 2.17.** *Theorem 2.16 also holds for the topologies  $\mathcal{T}_1(\mathbb{R}^2)$  and  $\mathcal{T}_1(\mathbb{R}^2)$ . Indeed if  $W(\mu_n, \mu)$  and  $W(\nu_n, \nu)$  converge to zero, Theorem 2.16 implies that  $\pi_n = \text{Curt}(\mu_n, \nu_n)$  converges to  $\pi = \text{Curt}(\mu, \nu)$  in  $\mathcal{T}_{\text{cb}}(\mathbb{R}^2)$ . Moreover we have  $\int |x| d\mu_n(x) \rightarrow \int |x| d\mu(x)$  and  $\int |y| d\nu_n(y) \rightarrow \int |y| d\nu(y)$ , so that  $\int (|x| + |y|) d\pi_n \rightarrow \int (|x| + |y|) d\pi$ . The convergence of the first moments and the convergence of  $\pi_n \rightarrow \pi$  in  $\mathcal{T}_{\text{cb}}(\mathbb{R}^2)$  imply  $W^{\mathbb{R}^2}(\pi_n, \pi) \rightarrow 0$  (see for instance [12, Proposition 7.12]).*

*Note that the quantitative statement established in Corollary 2.32 only provides the continuity for  $\mathcal{T}_1(\mathbb{R}^2)$  and  $\mathcal{T}_1(\mathbb{R}^2)$ . The proof relies on the fact that  $\text{Curt}(\mu, \nu)$  is a left-curtain coupling, while Theorem 2.16 uses that it is left-monotone.*

**Proof of Theorem 2.16.** Let us introduce a sequence  $(\mu_n, \nu_n)$  converging to  $(\mu, \nu)$ . We assume that for every  $n \in \mathbb{N}$ ,  $\pi_n$  is a left-monotone coupling of  $\mu_n$  and  $\nu_n$ . We will prove that  $\pi_n$  has a limit  $\pi$  and that it is also left-monotone. Due to Prokhorov’s theorem on compactness and the uniqueness of a left-monotone martingale coupling with given marginals, we can reduce the proof to the case we know that  $\pi_n$  converges to  $\pi$ .

We introduce the set  $E = \{(x^-, y^+, x^+, y^-, x', y') \in \mathbb{R}^6 | x^- \leq x^+ < x' \text{ and } y^- < y' < y^+\}$ . Assume that there is a vector  $v \in E \cap (\text{spt}^*(\pi))^3$ , which according to Corollary 2.15 is equivalent to the fact that  $\pi$  is not left-monotone. We will see that it implies that some  $\pi_n$  is not left-monotone. Before we proceed to the proof, let us stress that  $\pi_n^{\otimes 3}$  converges to  $\pi^{\otimes 3}$  and as  $v \in (\text{spt}(\pi^{\otimes 3})) = (\text{spt}(\pi))^3$  we obtain a sequence  $v_n$  with  $v_n \in (\text{spt}^*(\pi_n))^3$  and  $v_n \rightarrow v$ . Our goal will be to prove  $v_n \in E$  or directly that  $\pi_n$  is not left-monotone.

We distinguish two main cases.

*Case 1:  $v \in E^\circ$ .* As  $E^\circ$  is an open set,  $v_n \in E^\circ$  for  $n$  sufficiently large, which provides the contradiction with the fact that  $\pi_n$  is left-monotone.

*Case 2:  $v \in \partial E$ .* We have  $x^- = x^+$  and denote this real number simply by  $x$ . The arguments for the different subcases that we will distinguish are very similar to the ones in the proof of Proposition 2.14. The cases 2.1 and 2.2 corresponds to Case 1 and Case 2 of this proposition.

Recall that  $v_n = (x_n^-, y_n^+, x_n^+, y_n^-, x'_n, y'_n) \in \text{spt}^*(\pi_n)^3$  tends to  $v$ . If  $x_n^- \leq x_n^+$  we are done because with Corollary 2.15 this implies  $v_n \in E$ . Hence we must assume  $x_n^- > x_n^+$ .

*Case 2.1:  $y' \neq x$ .* This is not possible. If for instance  $x < y'$ , the relations  $x_n^- < x'_n$  and  $x_n^- < y'_n < y_n^+$  hold if  $n$  is sufficiently large so that we can use Lemma 2.13 for the points  $(x_n, y_n^+)$ ,  $(x'_n, y'_n) \in \text{spt}(\pi_n)$ . Hence this contradicts that  $\pi_n$  is left-monotone.

*Case 2.2.* Hence up to now we have assumed  $x = x^- = x^+$  and  $x_n^- > x_n^+$  and  $y' = x$ . We show now that  $x \in A(\pi)$ . Indeed if it is not true there exists a sequence  $(s_n, t_n) \in \text{spt}(\pi_n)$  converging to  $(s, t) \in \text{spt} \pi \cap ]-\infty, x[ \times ]x, +\infty[$ . Thus, if  $n$  is sufficiently large, recalling that  $y'_n$  tends to  $x$  we can apply Lemma 2.13 to  $(s_n, t_n)$  and  $(x'_n, y'_n)$ . Indeed we have  $s_t < x'_n$  and  $s_n < y'_n < t_n$ .

Let us see that  $x \in A \setminus A^-$  is impossible. Actually  $(x, y^+) \in \text{spt}^* \pi$  is an element of  $\text{spt} \pi$  and  $x < y^+$ . It follows  $\pi ]-\infty, x_1[ \times ]x_1, +\infty[ > 0$  for any  $x_1 \in ]x, y^+[$ . Hence  $x \in A^-$ . It follows  $\mu(x) > 0$ .

We assume for simplicity that  $x = 0$ . We denote  $\mu(x) \cdot \pi_x(]y^+/2, +\infty[)$  by  $m$ . It is not zero because  $y^+ \in \text{spt}(\pi_x)$ . Let  $\varepsilon > 0$  be strictly smaller than  $\min(x', y^+/2)$ . We also assume that it is sufficiently small to satisfy  $a = (y^+/2 - \varepsilon)(\pi_x(]y^+/2, +\infty[)/8) - \varepsilon > \varepsilon$  and  $\mu([- \varepsilon, \varepsilon]) < 2\mu(x)$ . We know that

$$\begin{aligned} & \liminf \pi_n(]1 - \varepsilon, +\varepsilon[ \times ]y^+/2, +\infty[) \\ & \geq \pi(]1 - \varepsilon, +\varepsilon[ \times ]y^+/2, +\infty[) \\ & \geq \mu(x) \cdot \pi_x(]y^+/2, +\infty[) = m, \end{aligned}$$

and for  $\mu_n = \text{proj}_{\#}^x \pi_n$ ,

$$\limsup \mu_n([- \varepsilon, \varepsilon]) \leq \mu([- \varepsilon, \varepsilon]).$$

Hence, there is  $n$  such that  $\pi_n(]1 - \varepsilon, +\varepsilon[ \times ]y^+/2, +\infty[) > m/2$  and  $\mu_n([- \varepsilon, \varepsilon]) < 2\mu([- \varepsilon, \varepsilon]) < 4\mu(x)$ . Therefore on a set  $B \subseteq ]- \varepsilon, \varepsilon[$  of positive  $\mu_n(ds)$ -measure the measure  $(\pi_n)_s(]y^+/2, +\infty[)$  is greater than  $\mu_n(]1 - \varepsilon, \varepsilon[)^{-1} \cdot m/2 > \pi_x(]y^+/2, +\infty[)/8$ . Hence for  $s \in B$ , using the fact that the barycenter of  $(\pi_n)_s$  is  $s$ , we have

$$(\pi_n)_s(]-\infty, -a[) > 0,$$

where  $a = (y^+/2 - s)(\pi_x(]y^+/2, +\infty[)/8) - s$ . Remind that  $-a < -\varepsilon < s$ . Let  $\Gamma$  be a Borelian set of  $\mathbb{R}^2$  such that  $\pi_n(\Gamma) = 1$ . As for almost every  $s$ , we have  $(\pi_n)_s(\{t \in \mathbb{R} \mid (s, t) \in \Gamma\}) > 0$ , we obtain that there are  $(s, t^-)$  and  $(s, t^+)$  in  $\Gamma$ , with  $t^+ > y^+/2$  and  $t^- < -\varepsilon$ , and  $(s', t') \in \Gamma$  close to  $(x', 0)$  such that  $t' \in ]t^-, t^+[$ . We conclude with Definition 2.7 that  $\pi_n$  can not be left-monotone, which contradicts our assumptions.  $\square$

**Remark 2.18.** *Theorem 2.16 provides a more direct and intuitive introduction of  $\pi_{1c} = \text{Curt}(\mu, \nu)$  than Definition 2.6. In this alternative presentation relying on [2, Section 2] (see also Lemma 2.33) one considers a sequence of atomic measures  $\mu_n$  that converges to  $\mu$  (see for instance point 3 in the proof of Proposition 2.34). We may assume  $\mu_n \leq_C \mu$  in order to satisfy  $(\mu_n, \nu) \in \mathcal{D}_{\leq_C}$ . The left-curtain couplings  $\pi_n = \text{Curt}(\mu_n, \nu)$  can be described easily, as is done for instance in the proof of Lemma 2.33. For that purpose it is not necessary to introduce the shadows in full generality but only to know what is the shadow of an atom. According to the theory  $\pi_{1c}$  is the limit of  $(\pi_n)_n$ .*

*Note that without the theory from [2], Theorem 2.16 can only prove that the accumulation points of the sequence  $(\pi_n)_n$  are all left-monotone couplings. Without Proposition 2.8 it is not known that the left-monotone elements of  $\Pi_M(\mu, \nu)$  are reduced to  $\{\pi_{1c}\}$ . Hence the alternative presentation explained in the present remark can not be seen as a definition.*

We end the paragraph on qualitative continuity with two results on the continuity of the shadows that will be useful in Section 2.3.

**Lemma 2.19 (Role of the mass of  $\nu$  close to  $\pm\infty$ ).** *Let  $\mu$  and  $\nu$  be measures of  $\mathcal{M}$  such that  $\mu \leq_{C,+} \nu$ . Let  $(\nu_n)_n$  such that  $\inf(\text{spt } \nu_n)$  tends to  $+\infty$ . Then the sequence  $S^{\nu+\nu_n}(\mu)$  tends to  $S^\nu(\mu)$  in  $\mathcal{M}$ .*

*The similar statements hold if  $\sup(\text{spt } \nu_n)$  tends to  $-\infty$  or  $\nu_n([a_n, b_n]) = 0$  with  $-a_n, b_n \rightarrow +\infty$ .*

**Proof.** 1. We will prove that the potential function of  $S^{\nu+\nu_n}(\mu)$  pointwise converges to the potential function of  $S^\nu(\mu)$ . Remind that it was introduced after  $W$  at the beginning of Section 1. Fix  $a \in \mathbb{R}$  and  $\varepsilon > 0$  and let  $\delta > 0$  be such that any measure  $\alpha \leq_+ \nu$  of mass  $\alpha(\mathbb{R}) \leq \delta$  satisfies  $\int |x - a| d\alpha \leq \varepsilon$ . Let  $\mu'$  be the leftmost measure smaller than  $\mu$  and of mass  $\mu'(\mathbb{R}) = \mu(\mathbb{R}) - \delta$ .

2. It is enough to prove that for any  $n$  satisfying  $\inf \text{spt } \nu_n > \sup[\text{spt } S^\nu(\mu')]$  we have

$$|u_{S^\nu(\mu)}(a) - u_{S^{\nu+\nu_n}(\mu)}(a)| = \left| \int |x - a| dS^\nu(\mu) - \int |x - a| dS^{\nu+\nu_n}(\mu) \right| \leq \varepsilon. \tag{3}$$

Before we state this inequality, we have to prove that  $\sup[\text{spt } S^\nu(\mu')]$  is finite. Actually, the shadow of  $\mu'$  restricted on  $] \sup \text{spt } \mu', +\infty[$  (more precisely  $(S^\nu(\mu') - \mu')_+|_{] \sup(\text{spt } \mu'), +\infty[}$ ) is made of the leftmost quantiles of  $\nu|_{\sup(\text{spt } \mu'), +\infty[}$  as is proved in Lemma 2.3. As some mass must remain for the shadow of  $\mu - \mu'$ , as explained in Proposition 2.4, this can not be the full  $\nu|_{\sup(\text{spt } \mu'), +\infty[}$ .

3. As  $S^v(\mu') \preceq_+ v + v_n$  and  $\mu' \preceq_C S^v(\mu')$ , we have  $S^{v+v_n}(\mu') \preceq_C S^v(\mu')$ . Considering Strassen's theorem (Proposition 1.5), we obtain the corollary that  $\sup(\text{spt } S^{v+v_n}(\mu')) \leq \sup(\text{spt } S^v(\mu'))$ . With the hypothesis made in 2 on the support of  $v_n$  this proves  $S^{v+v_n}(\mu') \preceq_+ v$ . Hence  $S^{v+v_n}(\mu') \succeq_C S^v(\mu')$ . Finally  $S^{v+v_n}(\mu') = S^v(\mu')$ .

4. We denote by  $\sigma$  the latter measure. Applying Proposition 2.4 to the shadow of the sum  $\mu' + (\mu - \mu')$  we get

$$\int |x - a| dS^{v+v_n}(\mu) = \int |x - a| d\sigma + \int |x - a| dS^{v+v_n-\sigma}(\mu - \mu')$$

and

$$\int |x - a| dS^v(\mu) = \int |x - a| d\sigma + \int |x - a| dS^{v-\sigma}(\mu - \mu').$$

As the shadow of  $\mu - \mu'$  in  $v + v_n - \sigma$  is smaller in the convex order than its shadow in  $v - \sigma$  and reminding the choices made in 1 we get

$$0 \leq \int |x - a| dS^{v+v_n-\sigma}(\mu - \mu') \leq \int |x - a| dS^{v-\sigma}(\mu - \mu') \leq \varepsilon,$$

so that (3) is established.

5. In the case  $\sup(\text{spt } v_n)$  tends to  $-\infty$  we just do the symmetric proof. If  $v_n([a_n, b_n]) = 0$  with  $-a_n, b_n \rightarrow +\infty$ , we implement a similar proof with the following modification: at step 1,  $\mu'$  is chosen in the middle of  $\mu$  so that its shadow has a compact support (adapt the argument of 2 that relies on Lemma 2.3). Steps 3 and 4 go in the same way. □

With Theorem 2.16 we obtain the corollary.

**Corollary 2.20.** *Under one of the three hypotheses of Lemma 2.19, we have  $\text{Curt}(\mu, S^{v+v_n}(\mu)) \xrightarrow{\mathcal{M}(\mathbb{R}^2)} \text{Curt}(\mu, S^v(\mu))$ .*

We remind another result of stability from [2].

**Proposition 2.21 (Proposition 4.15 in [2]).** *Assume that  $(\mu_n)_n$  is increasing in the convex order and  $\mu_n \preceq_C \mu \preceq_{C,+} v$  for every  $n \in \mathbb{N}$ . Then both  $(\mu_n)_n$  and  $(S^v(\mu_n))_n$  converge in  $\mathcal{M}$ . If we call  $\mu_\infty$ , respectively  $S_\infty$  the limits, then the measure  $S_\infty$  is the shadow of  $\mu_\infty$  in  $v$ .*

Again with Theorem 2.16 we obtain a corollary.

**Corollary 2.22.** *Under the hypotheses and notations of Proposition 2.21, we have  $\text{Curt}(\mu_n, S^v(\mu_n)) \xrightarrow{\mathcal{M}(\mathbb{R}^2)} \text{Curt}(\mu_\infty, S_\infty)$ .*

### 2.3. Lipschitz continuity of the shadow, quantitative estimates

In this section we give a quantitative version of Theorem 2.16 by using other methods. We start with definitions.

#### 2.3.1. Up and down measures

**Definition 2.23.** *Let  $\mu$  and  $\nu$  satisfy  $\mu(\mathbb{R}) = \nu(\mathbb{R})$  and call  $t$  this constant. We define the up and down measures of  $\mu$  and  $\nu$  as*

$$\text{Up}(\mu, \nu) = \max(G_\mu, G_\nu)_{\#\lambda_{[0,t]}} \quad \text{and} \quad \text{Down}(\mu, \nu) = \min(G_\mu, G_\nu)_{\#\lambda_{[0,t]}}.$$

#### Example 2.24.

- (i) *If  $\mu$  and  $\nu$  are probability measures and  $\text{Law}(X, Y)$  is the quantile coupling of these measures,  $\text{Up}(\mu, \nu)$  and  $\text{Down}(\mu, \nu)$  are simply the laws of  $\max(X, Y)$  and  $\min(X, Y)$ .*

(ii) If  $\mu = \sum_{i=1}^n \delta_{x_i}$  and  $\nu = \sum_{i=1}^n \delta_{y_i}$  with  $(x_i)_i, (y_i)_i$  non-decreasing sequences, then  $\text{Up}(\mu, \nu) = \sum_{i=1}^n \delta_{\max(x_i, y_i)}$  and  $\text{Down}(\mu, \nu) = \sum_{i=1}^n \delta_{\min(x_i, y_i)}$ .

**Lemma 2.25.** Let  $\mu$  and  $\nu$  be as in Definition 2.23. We have

$$\begin{aligned} W(\mu, \nu) &= W(\mu, \text{Down}(\mu, \nu)) + W(\text{Down}(\mu, \nu), \nu) \\ &= W(\mu, \text{Up}(\mu, \nu)) + W(\text{Up}(\mu, \nu), \nu). \end{aligned}$$

**Proof.** The proof simply relies on the formula  $W(\mu, \nu) = \int |G_\nu - G_\mu| d\lambda$  in Lemma 1.4 and  $|G_\nu - G_\mu| = (G_\mu - G) + (G_\nu - G) = (G' - G_\mu) + (G' - G_\nu)$  where  $G = \min(G_\mu, G_\nu)$  and  $G' = \max(G_\mu, G_\nu)$ .  $\square$

**Lemma 2.26.**  $\text{Up}(\mu, \mu') + \nu = \text{Up}(\mu + \nu, \mu' + \nu)$ .

**Proof.** One can check with the definition of Up by random variables that  $F_{\text{Up}(\mu, \mu')}(t) = \min(F_\mu(t), F_{\mu'}(t))$ . Hence for every  $t \in \mathbb{R}$

$$\begin{aligned} F_{\text{Up}(\mu, \mu') + \nu}(t) &= \min(F_\mu(t), F_{\mu'}(t)) + F_\nu(t) \\ &= \min(F_\mu(t) + F_\nu(t), F_{\mu'}(t) + F_\nu(t)). \end{aligned}$$

But this is also the cumulative distribution function of  $\text{Up}(\mu + \nu, \mu' + \nu)$ . If  $\mu, \mu'$  are not probability measures, the result follows from their normalisation.  $\square$

The shadow  $S^\nu(\mu)$  is only defined for  $\mu \preceq_{C,+} \nu$ , which may be restrictive for some proofs. But if, roughly speaking, we add mass close to infinity to  $\nu$  this shadow can exist. The two next lemmas permit us to implement this idea, which for instance plays a role in the final proof of Section 2.

**Lemma 2.27 (Adding mass at  $\pm\infty$ ).** If  $\mu \preceq_{C,+,\text{sto}} \nu$ , for any  $a \in \mathbb{R}$  there exists  $\nu'$  with  $\nu'$  concentrated on  $] -\infty, a]$  and  $\mu \preceq_{C,+} \nu + \nu'$ .

Similarly, if no assumption is done on  $\mu$ , for any  $a, b \in \mathbb{R}$ , there exists  $\nu'$  with  $\nu'(\{a, b\}) = 0$  and  $\mu \preceq_{C,+} \nu'$ .

**Proof.** 1. We start to prove the first result in the special case  $\mu \preceq_{+,\text{sto}} \nu$ . For any  $a \in \mathbb{R}$ , we will find  $\nu'$  with  $\nu'(\{a, +\infty\}) = 0$  and  $\mu' \preceq_{C,+} \nu + \nu'$ . Let assume without loss of generality that  $\mu$  is a probability measure and applying Theorem 1.7 let  $X \leq Y$  be random variables with laws  $\mu = \text{Law}(X)$  and  $\text{Law}(Y) \preceq_+ \nu$ . Let us fix  $a \in \mathbb{R}$  and consider in a first time the case  $\mu(\{ -\infty, a\}) = 0$ . In this case we introduce  $U$  a random variable independent from  $X$  and define  $Z$

$$Z = \begin{cases} Y & \text{if } U \geq \frac{Y-X}{Y-a}, \\ a & \text{otherwise.} \end{cases}$$

Therefore  $\mathbb{E}(Z | X) = X$ . Observe that  $\text{Law}(Z) \preceq_+ \nu + \mathbb{P}(Z = a)\delta_a$  so that for  $\nu' = \delta_a$  we have  $\mu \preceq_{C,+} \nu + \nu'$ .

In the second case we can write  $\mu = \mu_1 + \mu_2$  where  $\mu_1$  is concentrated on  $A = ] -\infty, a]$  and  $\mu_2(A) = 0$ . By using the result just proved for  $\mu_2$ , we obtain  $\nu'_2$  concentrated on  $A$  such that  $\mu_2 \preceq_{C,+} \nu + \nu'_2$ . We set  $\nu'_1 = \mu_1$  so that we also have  $\mu_1 \preceq_{C,+} \nu'_1$ . As  $\nu' = \nu'_1 + \nu'_2$  is concentrated on  $A$ , we are done.

2. If  $\mu \preceq_{C,+,\text{sto}} \nu$ , according to Theorem 1.7 there exists  $\mu'$  such that  $\mu \preceq_C \mu'$  and  $\mu' \preceq_{+,\text{sto}} \nu$ . Therefore with paragraph 1, we find for every  $a \in \mathbb{R}$  a measure  $\nu'$  concentrated on  $] -\infty, a]$  such that  $\mu' \preceq_{C,+} \nu + \nu'$ . But  $\mu \preceq_{C,+} \mu'$  so that  $\mu \preceq_{C,+} \nu + \nu'$  is also satisfied.

3. For the second statement fix  $a$  and  $b$  and set  $\nu = \mu(\{ -\infty, b\})\delta_b + \mu|_{]b, \infty[}$ . As  $\mu \preceq_{\text{sto}} \nu$ , one can apply part 1. of the present proof.  $\square$

### 2.3.2. Semimetric $Z$ on the space of transport plans

A semimetric is a function that is positive-definite and symmetric. In other words it satisfies all axioms of a distance except the triangle inequality.

**Definition 2.28.** Let  $\pi$  and  $\pi'$  be two transport plans. We define  $Z(\pi, \pi')$  as

$$\begin{aligned} Z(\pi, \pi') &= \inf_{(\pi^s, \pi'^s)_{s \in [0,1]}} \sup_{s \in [0,1]} \max(W(\mu^s, \mu'^s), W(\nu^s, \nu'^s)) \\ &= \max\left(W(\mu, \mu'); \inf_{(\pi^s, \pi'^s)_{s \in [0,1]}} \sup_{s \in [0,1]} W(\nu^s, \nu'^s)\right), \end{aligned}$$

where  $\mu^s, \nu^s$  are the marginals of  $\pi^s$  and  $\mu'^s, \nu'^s$  the marginals of  $\pi'^s$ . The infimum is taken among all the families  $(\pi^s)_{s \in [0,1]}, (\pi'^s)_{s \in [0,1]}$  that satisfy

1.  $\forall s \in [0, 1], \pi^s(\mathbb{R}^2) = s$ ,
2.  $\forall s \in [0, 1], \exists x \in [-\infty, +\infty], \pi_{]-\infty, x[ \times \mathbb{R}} \leq + \pi^s \leq + \pi_{]-\infty, x[ \times \mathbb{R}}$  (in fact  $x = G_\mu(s)$ ),
3. if  $s \leq t$ , then  $\pi^s \leq + \pi^t$ .

In the *Main theorem* we use the most natural extension of  $Z$  to the measures of mass  $\lambda > 0$  given by  $Z(\lambda\pi, \lambda\pi') = \lambda Z(\pi, \pi')$ .

**Remark 2.29.** If the first marginals of  $\pi$  and  $\pi'$  are continuous, there is no freedom in the choice of  $(\pi^s)_{s \in [0,1]}, (\pi'^s)_{s \in [0,1]}$ . Hence  $Z$  can be reformulated as

$$Z(\pi, \pi') = \max \left\{ \begin{array}{l} W(\mu, \mu'), \\ \sup_{F_\mu(x) = F_{\mu'}(x')} W(\text{proj}_\#^y \pi |_{]-\infty, x[ \times \mathbb{R}}, \text{proj}_\#^y \pi' |_{]-\infty, x'[ \times \mathbb{R}}). \end{array} \right.$$

**Proposition 2.30.** The function  $Z$  is a semimetric on  $\mathcal{P}(\mathbb{R}^2)$  and the triangle inequality is satisfied on the subspace of measures with continuous first marginal. Moreover if  $Z(\pi_n, \pi) \rightarrow 0$  we have  $\pi_n \rightarrow \pi$  for the topology  $\mathcal{T}_1(\mathbb{R}^2)$ .

**Proof.** 1. It is clear that  $Z$  is symmetric and  $Z(\pi, \pi') = 0$  if and only if  $\pi = \pi'$ . This principle is used in the definition of the left-curtain coupling (see Definition 2.6): the measures  $\text{proj}_\#^y \pi |_{]-\infty, x[ \times \mathbb{R}}$  completely determine  $\pi$ . The triangle inequality on the subspace of measures with continuous first marginal follows from the triangle inequality of  $W$  and Remark 2.29.

2. Assume that  $Z(\pi_n, \pi)$  tends to 0. For every  $n$ , introduce families  $(\pi_n^s)_{s \in [0,1]}$  that together with some families decomposing  $\pi$  satisfy the limit condition. We introduce a sequence  $\pi_n'$  of transport plans defined as intermediate measure between the measures  $\pi_n$  and  $\pi$ . Their first marginals are  $\text{proj}_\#^x \pi_n' = \mu$  like  $\pi$  and one associates them with the same family  $\nu_n^s = \text{proj}_\#^y \pi_n^s$  like  $\pi_n$ . More precisely, let  $(U, X, (X_n, Y_n))_{n \in \mathbb{N}}$  be such that

- $\text{Law}(U) = \lambda_{[0,1]}$ ,
- $(U, X, X_n)$  is comonotonic (i.e.  $X = G_X(U)$  and  $X_n = G_{X_n}(U)$  almost surely),
- $\text{Law}(X) = \mu$ ,
- $\text{Law}(X_n, Y_n) = \pi_n$ ,
- $\nu_n^s = s \text{Law}(Y_n | U \leq s)$ ,

then  $\pi_n'$  is the law of  $(X, Y_n)$  and  $\pi_n = \text{Law}(X_n, Y_n)$ .

For any  $s \in [0, 1]$ , let  $\nu^s$  be  $\text{proj}_\#^y \pi^s$  with  $(\pi^s)_{s \in [0,1]}$  some family admissible in the sense of Definition 2.28(1)–(3). We do not necessarily assume  $W(\nu_n^s, \nu^s) \rightarrow 0$  for every  $s$  but we will need the following remark in our proof: for any  $x \in \mathbb{R}$ , the measure  $\nu^{F_\mu(x)}$  is completely characterised and it is  $\text{proj}_\#^y (\pi |_{]-\infty, F_\mu(x)[ \times \mathbb{R}})$ . Indeed this is the only measure of mass  $F_\mu(x)$  satisfying point (2) in Definition 2.28. Hence for any  $x \in \mathbb{R}$ ,  $W(\nu_n^{F_\mu(x)}, \nu^{F_\mu(x)}) \rightarrow 0$  because the family  $(\nu_n^s)_{s \in [0,1]}$  corresponding to  $(\pi_n^s)_{s \in [0,1]}$  has been properly chosen for the convergence.

We can now proceed to the proof. We want to prove  $W^{\mathbb{R}^2}(\pi_n, \pi_n') \rightarrow 0$  and  $W^{\mathbb{R}^2}(\pi_n', \pi) \rightarrow 0$ . On the one hand we have  $W^{\mathbb{R}^2}(\pi_n, \pi_n') \leq \mathbb{E}(|X - X_n| + |Y_n - Y_n|) = W(\mu_n, \mu) \leq Z(\pi_n, \pi) \rightarrow 0$ . On the other hand, we prove that at any continuity point  $(x, y) \in \mathbb{R}^2$  of  $F_\pi$ , the sequence  $F_{\pi_n}$  pointwise converges to  $F_\pi$ . According to a classical characterisation (see e.g. [3, Example 2.3]), this will imply  $\pi_n \rightarrow \pi$  in the weak topology  $\mathcal{T}_{\text{cb}}(\mathbb{R}^2)$ . Moreover, as  $W(\nu, \nu_n) = W(\nu^1, \nu_n^1) \rightarrow 0$  and  $\mu$  is the first marginal of both  $\pi_n'$  and  $\pi$ , one can apply Remark 2.17. Therefore the weak convergence  $\pi_n' \rightarrow \pi$  implies the weak convergence with finite first moments  $W^{\mathbb{R}^2}(\pi_n', \pi) \rightarrow 0$ .



Let  $(x, y)$  be a continuity point of  $F_\pi$ . We have  $F_\pi(x, y) = \nu^{F_\mu(x)}(-\infty, y]$  and also  $F_{\pi'_n}(x, y) = \pi'_n(-\infty, x] \times ]-\infty, y] = \nu_n^{F_\mu(x)}(-\infty, y]$ . This tends to  $\nu^{F_\mu(x)}(-\infty, y]$  because  $W(\nu^{F_\mu(x)}, \nu_n^{F_\mu(x)}) \rightarrow 0$  and  $y$  is a continuity point of  $\nu^{F_\mu(x)}$ . We conclude that  $\pi'_n$  weakly converges to  $\pi$ . Therefore

$$W^{\mathbb{R}^2}(\pi_n, \pi) \leq W^{\mathbb{R}^2}(\pi_n, \pi'_n) + W^{\mathbb{R}^2}(\pi'_n, \pi) \rightarrow 0. \quad \square$$

The following estimate is one of our main theorems. It provides a quantitative estimate on the Lipschitz continuity of the shadow projection  $(\mu, \nu) \mapsto S^\nu(\mu)$ .

**Theorem 2.31.** *Let  $\mu, \mu'$  and  $\nu, \nu'$  be elements  $\mathcal{M}$ . We assume  $\mu \leq_{C,+} \nu$  and  $\mu' \leq_{C,+} \nu'$  respectively. We assume also  $\mu(\mathbb{R}) = \mu'(\mathbb{R})$  and  $\nu(\mathbb{R}) = \nu'(\mathbb{R})$ . The following relation holds*

$$W(S^\nu(\mu), S^{\nu'}(\mu')) \leq W(\mu, \mu') + 2W(\nu, \nu').$$

The proof of the theorem is postponed at the end of the section. It relies on all results in between including Proposition 2.34 and Proposition 2.36. The first proposition is concerned with  $\nu = \nu'$  and the second with  $\mu = \mu'$ . Before we start with this program let us state a corollary of Theorem 2.31 that gives a quantitative turn to Theorem 2.16 in terms of the semimetric  $Z$ . A similar result can not be satisfied with  $W^{\mathbb{R}^2}$  in place of  $Z$  as is explained later in Example 2.37.

**Corollary 2.32.** *Consider the mapping  $\text{Curt} : (\mu, \nu) \in \mathcal{D}_{\leq C} \mapsto \pi_{1c}$ , where  $\mathcal{D}_{\leq C} = \{(\mu, \nu) \in \mathcal{P}^2 : \mu \leq_C \nu\}$ . This mapping is continuous from  $\mathcal{D}_{\leq C} \subseteq \mathcal{P}^2$  to  $\Pi_M \subseteq \mathcal{P}(\mathbb{R}^2)$  equipped with the topologies  $\mathcal{T}_1(\mathbb{R}^2)^2$  and  $\mathcal{T}_1(\mathbb{R}^2)$ . More precisely*

$$Z(\text{Curt}(\mu, \nu), \text{Curt}(\mu', \nu')) \leq W(\mu, \mu') + 2W(\nu, \nu').$$

**Proof.** The continuity is obtained from the estimate in the statement and Proposition 2.30. Let us prove the estimate. Like in paragraph 1.2 we have  $\mu = (G_\mu)_{\# \lambda_{[0,1]}}$ . We introduce  $\mu^s = (G_\mu)_{\# \lambda_{[0,s]}}$  and  $\nu^s = S^\nu(\mu^s)$ . In a similar way to Definition 2.6 of  $\pi_{1c}$ , we obtain a unique family  $\pi^s$ , increasing for  $\leq_+$ , and with marginals  $\mu^s$  and  $\nu^s$ . We proceed in the same way for  $\mu'$  and  $\nu'$ . We obtain the wanted estimate by applying Theorem 2.31 to these measures.  $\square$

We start with the preliminaries of the proof of Theorem 2.31.

### 2.3.3. Variations in $\mu$

**Lemma 2.33 (Important lemma).** *Let  $n$  an integer and  $\mu, \mu'$  be two measures that are the sum of  $n$  atoms of the same mass and such that  $\mu \leq_{\text{sto}} \mu'$ . If  $\nu \in \mathcal{M}$  satisfies  $\mu \leq_{C,+} \nu$  and  $\mu' \leq_{C,+} \nu$  we have  $S^\nu(\mu) \leq_{\text{sto}} S^\nu(\mu')$ .*

**Proof.** Without loss of generality, we can assume that the atoms are all Dirac masses (the mass is 1). We write  $\mu = \sum_{i=1}^n \delta_{x_i}$  with  $x_i \leq x_{i+1}$  for any  $i < n$  and use the same notations for  $\mu'$ . As  $\mu \leq_{\text{sto}} \mu'$ , one has  $x_i \leq x'_i$  for any  $i \leq n$ .

The proof relies on the description of the shadow of a measure concentrated in one point as  $G_{\#\lambda_{]q, q+\alpha]}$  where  $G$  is the inverse cumulative function of the target measure  $G_\nu$  and  $\alpha$  is the mass of the atom (see Example 2.2). It relies also on the decomposition Proposition 2.4 as in Example 2.5: if  $\bar{\mu}$  (resp.  $\bar{\mu}'$ ) is the restriction of  $\mu$  (resp.  $\mu'$ ) to  $\{x_1, \dots, x_{n-1}\}$  (resp.  $\{x'_1, \dots, x'_{n-1}\}$ ) we have

$$S^\nu(\mu) = S^\nu(\bar{\mu}) + S^{\nu - S^\nu(\bar{\mu})}(\delta_{x_n})$$

and the similar equation holds for  $\mu', \bar{\mu}'$  and  $\delta_{x'_n}$ .

We will prove the result by induction on  $n$ , the number of atoms, not greater than  $m = \nu(\mathbb{R})$ . For  $n = 1$ , the statement is obvious. Actually, denoting by  $G$  the inverse cumulative function of  $\nu$  (it satisfies  $G_{\#\lambda_{]0, m]} = \nu$ ) the shadow measures can be written as  $G_{\#\lambda_{]p, p+1]}$  and  $G_{\#\lambda_{]p', p'+1]}$  where  $p \leq p'$ .

If  $n \geq 2$  we adopt the notations  $\bar{\mu}, \bar{\mu}'$  introduced above and we assume a statement stronger than the lemma that we call  $\mathcal{H}_{n-1}$ : there exists two sets  $\bar{J}, \bar{J}' \subseteq ]0, m]$  that satisfy

- the masses of  $\bar{J}$  and  $\bar{J}'$  are  $n - 1$ ,
- $\bar{J}$  is a disjoint union of intervals of type  $]a, b]$ . The same holds for  $\bar{J}'$ ,
- these intervals have integer length,
- $\lambda_{\bar{J}} \preceq_{\text{sto}} \lambda_{\bar{J}'}$  (during this proof we note it simply  $\bar{J} \preceq_{\text{sto}} \bar{J}'$ ). In particular,  $\max(\bar{J}) \leq \max(\bar{J}')$ ,
- $G_{\#} \lambda_{\bar{J}} = S^v(\bar{\mu})$  and  $G_{\#} \lambda_{\bar{J}'} = S^v(\bar{\mu}')$  (in particular  $S^v(\bar{\mu}) \preceq_{\text{sto}} S^v(\bar{\mu}')$ ).

As  $v$  may possess atoms of mass  $> 1$ , the sets  $\bar{J}$  and  $\bar{J}'$  may not be unique. We assume that  $\max \bar{J}$  and  $\max \bar{J}'$  are as small as possible: no other set satisfying the five conditions before has a smaller maximum. Note also that we proved  $\mathcal{H}_1$  in the paragraph above.

Starting from  $\bar{J}$  and  $\bar{J}'$  as in  $\mathcal{H}_{n-1}$  where  $n \leq m$ , we now construct sets  $J \supseteq \bar{J}$  and  $J' \supseteq \bar{J}'$  that satisfy  $\mathcal{H}_n$  where  $\mu, \mu'$  replace  $\bar{\mu}, \bar{\mu}'$ . In fact we follow the way described at the beginning of the proof and look for the shadow of  $\delta_{x_n}$  (resp.  $\delta_{x'_n}$ ) in  $v - S^v(\bar{\mu}) = G_{\#} \lambda_{]0, m[ \setminus \bar{J}}$  (resp.  $v - S^v(\bar{\mu}')$ ). We obtain the restriction of these measures to a “quantile interval,” which can be described as  $G_{\#} \lambda_{I \setminus \bar{J}}$  and  $G_{\#} \lambda_{I' \setminus \bar{J}'}$ , where  $I = ]p, q]$  and  $I' = ]p', q']$  are intervals and  $I \setminus \bar{J}$  and  $I' \setminus \bar{J}'$  are pseudo-intervals (see Example 2.5) of Lebesgue measure 1. If the shadow is a Dirac mass, several choices of  $I$  may be available. We choose the smallest possible  $\max I$ , respectively  $\max I'$ . Finally the sets  $J = \bar{J} \cup I$  and  $J'$  are the union of intervals of integer length. Moreover  $S^v(\mu) = G_{\#} \lambda_J$  and  $S^v(\mu') = G_{\#} \lambda_{J'}$  as we want.

We still have to prove the relation  $J \preceq_{\text{sto}} J'$ . Our first step is to see  $\max I > \max \bar{J}$  (and  $\max I' > \max \bar{J}'$ , which can be shown in the same way). Indeed it is clear if  $x_n > x_{n-1}$ . If  $x_n = x_{n-1}$ , a problem may happen if the shadows of both  $\delta_{x_{n-1}}$  and  $\delta_{x_n}$  are  $\delta_{x_n}$ , but we recall that  $\max \bar{J}$  was the smallest possible value coherent with  $\mathcal{H}_{n-1}$  so that  $\max I > \max \bar{J}$  as we want. Therefore  $q := \max I$  and  $\bar{J}$  completely determine  $J$ . Indeed, one obtains  $J$  in adding the greatest real numbers in  $]0, m] \setminus \bar{J}$  that are smaller than  $q$ . One proceeds until the set has measure 1. The result can also be written  $J = \bar{J} \cup ]p, q]$ .

One can see the barycenter of  $\mu$  as a continuous and increasing function of  $q$  and the same is true for  $\mu'$  and  $q' = \sup J'$ . Let us fix  $q'$  and consider  $q$  as a variable. As  $\bar{J} \preceq_{\text{sto}} \bar{J}'$ , one has also  $(]0, m] \setminus \bar{J}') \preceq_{\text{sto}} (]0, m] \setminus \bar{J})$ . Hence if for the shadows of  $x_n$  and  $x'_n$  in  $v - S^v(\bar{\mu})$  and  $v - S^v(\bar{\mu}')$  respectively, we start from the same value  $q = q' > \max \bar{J}' \geq \max \bar{J}$  at the right of the interval  $]0, m]$  and collect to the left the real numbers in  $]0, m] \setminus \bar{J}$  and  $]0, m] \setminus \bar{J}'$  respectively until one has a set of mass 1, the set that we obtain for  $S^{v-S^v(\mu)}(\delta_{x_n})$  is stochastically greater as the one for  $S^{v-S^v(\mu')}(\delta_{x'_n})$ . In other words  $I' \setminus \bar{J}' \preceq_{\text{sto}} I \setminus \bar{J}$ . This relation still holds after the push-forward  $G_{\#}$ . Taking the barycenters, we obtain  $x'_n \leq x_n$ . But the hypothesis of the lemma states  $x_n \geq x'_n$ . Having in mind the continuity and the monotonicity of  $x_n$  with respect to  $q$  we see that the correct position of  $q$  satisfies  $q \leq q'$ .

The length of the rightmost interval of  $J'$  is an integer that we denote by  $k$ . As  $q' \geq q$  the upper part of mass  $k$  of  $S^v(\mu')$  in the stochastic order is greater, for the same order, than the corresponding measure part of  $S^v(\mu)$ . The rest of  $J'$  is included in  $\bar{J}'$ . Due to the induction, it is greater than the corresponding left part of  $\bar{J}$  of mass  $n - k$ . This left part of  $\bar{J}$  is greater (in the stochastic order) than the left part of  $J$ , that is the most left part of mass  $n - k$ , because  $\bar{J} \subseteq J$ . Thus  $J \preceq_{\text{sto}} J'$  and this pair fulfills  $\mathcal{H}_n$ . □

**Proposition 2.34.** *Let  $\mu, \mu' \in \mathcal{M}$  with the same mass. Assume  $v \in \mathcal{M}$  satisfies  $\mu \preceq_{C,+} v$  and  $\mu' \preceq_{C,+} v$ . We have*

$$W(S^v(\mu), S^v(\mu')) \leq W(\mu, \mu'). \tag{4}$$

**Proof.** 1. We first assume that  $\mu$  and  $\mu'$  are made of finitely many atoms of the same mass. We also assume  $\text{Down}(\mu, \mu') \preceq_{C,+} v$  and we denote this measure by  $\tilde{\mu}$ . As explained in Example 2.24,  $\tilde{\mu}$  is a measure of the same type as  $\mu$  and  $\mu'$ . Hence we can apply Lemma 2.33 to the pairs  $(\mu, \tilde{\mu})$  and  $(\mu', \tilde{\mu})$ . Using Lemma 1.4 and Lemma 2.25 we can compute as follows

$$\begin{aligned} W(\mu, \mu') &= W(\mu, \tilde{\mu}) + W(\mu', \tilde{\mu}) \\ &= W(S^v(\mu), S^v(\tilde{\mu})) + W(S^v(\mu'), S^v(\tilde{\mu})) \\ &\geq W(S^v(\mu), S^v(\mu')). \end{aligned}$$

2. If  $\tilde{\mu} \preceq_{C,+} v$  does not hold, we have  $\tilde{\mu} \preceq_{C,+,\text{sto}} v$  so that there exists  $(v_n)_n$  as in Lemma 2.27. We have  $\tilde{\mu} \preceq_{C,+} v + v_n$  where  $\text{sup spt } v_n$  tends to  $-\infty$  and the computation above leads to  $W(S^{v+v_n}(\mu), S^{v+v_n}(\mu')) \leq W(\mu, \mu')$ . Therefore with Lemma 2.19, we obtain (4).

3. For general  $\mu, \mu'$  of the same mass  $m$ , that we assume to be 1, we approach them in  $\mathcal{P}$  by measures  $\mu_n \leq_C \mu$  and  $\mu'_n \leq_C \mu'$  with the same barycenter, obtained as the sum of  $2^n$  atoms of mass  $m/2^n$ . We do it in the following way:  $\mu_n$  is defined as  $\sum_{k=1}^{2^n} \frac{1}{2^n} \delta_{x_k}$  where

$$x_k = 2^n \int_{k/2^n}^{(k+1)/2^n} G_\mu(t) \, d\lambda(t).$$

The quantile function associated with  $\mu_n$  is constant on each  $[k/2^n, (k+1)/2^n]$  with value the mean of  $G_\mu$  on this interval. We recognise for the filtration made of the dyadic intervals of  $]0, 1]$ , the martingale associated with the random variable  $G_\mu \in L^1(]0, 1])$ , the  $L^1$ -norm being the Kantorovich distance between the measures, as explained in Lemma 1.4. Hence  $(\mu_n)_n$  is non decreasing for the convex order and  $\mu_n \rightarrow \mu$  in  $\mathcal{M}$ . Thus applying Proposition 2.21 we obtain the wanted estimate as  $n$  goes to  $\infty$ .  $\square$

**Remark 2.35.** One can relax the assumption to have atomic measures in Lemma 2.33 by using the approximation detailed in point 3 of the proof of Proposition 2.34. Indeed, the stochastic order is stable in the weak topology so that  $S^v(\mu) \leq_{\text{sto}} S^v(\mu')$  is true for general measures  $\mu \leq_{\text{sto}} \mu'$ .

### 2.3.4. Variations in $v$ and conclusion

**Proposition 2.36.** Let  $\mu, v$  and  $v'$  be elements of  $\mathcal{M}$  such that  $\mu \leq_{C,+} v, \mu \leq_{C,+} v'$  and  $v(\mathbb{R}) = v'(\mathbb{R})$ . Then it holds

$$W(S^v(\mu), S^{v'}(\mu)) \leq 2W(v, v').$$

**Proof.** 1. We make first the additional assumption  $v \leq_{\text{sto}} v'$  and we will prove  $W(S^v(\mu), S^{v'}(\mu)) \leq 2W(v, v')$  in this case. Because of Proposition 2.21, we can assume without loss of generality that  $\mu$  is of type  $\sum_{i=1}^n \alpha_i \delta_{x_i}$  by using the same method as at step 3 of Proposition 2.34. We can describe  $S^v(\mu)$  as it is done in Example 2.5 and introduce for this purpose a sequence  $J_1 \subseteq \dots \subseteq J_n$ . We have  $S^v(\mu) = (G_v)_{\#} \lambda_{J_n}$  and for any  $k, S^v(\sum_{i=1}^k \alpha_i \delta_{x_i}) = (G_v)_{\#} \lambda_{J_k}$ . We introduce now  $S' = (G_{v'})_{\#} \lambda_{J_n}$  and  $\mu' = \sum_{i=1}^n \alpha_i \delta_{x'_i}$  where  $x'_i$  is the barycenter of  $(G_{v'})_{\#} \lambda_{J_i \setminus J_{i-1}}$ . As  $v \leq_{\text{sto}} v'$ , we have  $G_v \leq G_{v'}$  and  $S^v(\mu) \leq_{\text{sto}} S'$ . Of course  $x_i \leq x'_i$  so that  $\mu \leq_{\text{sto}} \mu'$ . According to the converse statement in Example 2.5 we also have  $S' = S^{v'}(\mu')$ . Therefore using Proposition 2.34 for  $\mu, \mu' \leq_{C,+} v'$  we obtain

$$\begin{aligned} W(S^v(\mu), S^{v'}(\mu)) &\leq W(S^v(\mu), S^{v'}(\mu')) + W(S^{v'}(\mu'), S^{v'}(\mu)) \\ &\leq W(v, v') + W(\mu, \mu') \leq 2W(v, v'). \end{aligned}$$

Indeed, due to  $\mu \leq_{\text{sto}} \mu'$  and  $S^v(\mu) \leq_{\text{sto}} S^{v'}(\mu')$  we have

$$W(\mu, \mu') = W(S^v(\mu), S^{v'}(\mu')) = \int_{J_n} (G_{v'} - G_v) \, d\lambda \leq \int_0^{v(\mathbb{R})} (G_{v'} - G_v) \, d\lambda = W(v, v').$$

2. We assume now  $\mu \leq_{C,+} \text{Up}(v, v')$ . In this case we use the triangle inequality, paragraph 1 and Lemma 2.25 so that we can establish

$$W(S^v(\mu), S^{v'}(\mu)) \leq 2W(v, v').$$

3. Let us assume that  $\mu \leq_{C,+} \text{Up}(v, v')$  does not hold. According to Lemma 2.27, there exists  $\gamma$  such that  $\mu \leq_{C,+} \text{Up}(v, v') + \gamma$ . But as stated in Lemma 2.26,  $\text{Up}(v, v') + \gamma$  is also  $\text{Up}(v + \gamma, v' + \gamma)$  and  $\mu \leq_{C,+} v + \gamma$  as well as  $\mu \leq_{C,+} v' + \gamma$ . Therefore according to the previous paragraph

$$W(S^{v+\gamma}(\mu), S^{v'+\gamma}(\mu)) \leq 2W(\mu, \mu').$$

Note that in Lemma 2.27,  $\text{sup}(\text{spt}(\gamma))$ , that is the upper bound on the support of  $\gamma$  can be chosen arbitrary close to  $-\infty$ . Hence, letting  $\text{sup}(\text{spt}(\gamma))$  go to  $-\infty$ , Lemma 2.19 permits us to conclude in the most general case.  $\square$

**Proof of Theorem 2.31.** We combine Proposition 2.34 and Proposition 2.36 and simply use the triangle inequality

$$\begin{aligned} W(S^v(\mu), S^{v'}(\mu')) &\leq W(S^v(\mu), S^{v'}(\mu)) + W(S^{v'}(\mu), S^{v'}(\mu')) \\ &\leq 2W(v, v') + W(\mu, \mu'). \end{aligned}$$

This can only be written doing the further assumption  $\mu \preceq_{C,+} v'$ .

In the general case, let  $\gamma$  be such that  $\mu \preceq_{C,+} v' + \gamma$ . Of course we still have  $\mu \preceq_{C,+} v + \gamma$  and  $\mu' \preceq_{C,+} v' + \gamma$ . The previous computation holds and writes

$$W(S^{v+\gamma}(\mu), S^{v'+\gamma}(\mu')) \leq 2W(v + \gamma, v' + \gamma) + W(\mu, \mu').$$

But  $W(v + \gamma, v' + \gamma) = W(v, v')$ . We conclude using the same method as at the end of the proof of Proposition 2.36. With the end of Lemma 2.27 we obtain a suitable sequence  $(\gamma_n)_n$  and we use the end of Lemma 2.19 for the convergence  $W(S^{v+\gamma_n}(\mu), S^{v'+\gamma_n}(\mu')) \rightarrow W(S^v(\mu), S^{v'}(\mu'))$ .  $\square$

The following example shows that the left-curtain Curt is not Lipschitzian when considering the Kantorovich distances  $W$  and  $W^{\mathbb{R}^2}$ . In other words a result like Corollary 2.32 does not hold.

**Example 2.37.** For any integer  $n \geq 1$  and  $\varepsilon < 1$ , we consider four measures of mass  $n$ . The first two involve Dirac masses at some points  $k \in \mathbb{N}$  and  $\varepsilon > 0$ .

$$\mu_n = \sum_{k=0}^{n-1} \delta_k \quad \text{and} \quad \mu'_n = \delta_\varepsilon + \sum_{k=1}^{n-1} \delta_k$$

and the two others are made of uniform measures

$$v_n = \lambda_{[-1/2, n-1/2]} \quad \text{and} \quad v'_n = \lambda_{[-1/2+\varepsilon/n, n-1/2+\varepsilon/n]}.$$

Note that  $\mu_n \preceq_{\text{sto}} \mu'_n$ ,  $v_n \preceq_{\text{sto}} v'_n$  and  $W(\mu_n, \mu'_n) = W(v_n, v'_n) = \varepsilon$ . As the measures are pairwise in convex order we can define the curtain couplings  $\pi_n = \text{Curt}(\mu_n, v_n)$  and  $\pi'_n = \text{Curt}(\mu'_n, v'_n)$ . We have

$$\pi_n = \sum_{k=0}^{n-1} \delta_k \otimes \lambda_{[k-1/2, k+1/2]}.$$

The expression of  $\pi'_n$  is more intricate.

$$\pi'_n = \delta_\varepsilon \otimes \lambda_{[-1/2+\varepsilon, 1/2+\varepsilon]} + \sum_{k=1}^{n-1} \delta_k \otimes \lambda_{A_{k,1} \cup A_{k,2}},$$

where for any  $k \leq n - 1$ , the set  $A_{k,1} \cup A_{k,2} = [-1/2 + \varepsilon/(k + 1), -1/2 + \varepsilon/k] \cup [k - 1/2 + \varepsilon/k, k + 1/2 + \varepsilon/k + 1]$  is the union of an interval of length  $\frac{\varepsilon}{k(k+1)}$  close to  $-1/2$  and the interval of length  $1 - \frac{\varepsilon}{k(k+1)}$  with barycenter close to  $k$ .

In  $\mathbb{R}^2$  the set  $\{k\} \times A_{k,1}$  is part of the support of  $\pi'_n$ . It has mass  $\varepsilon/k(k + 1)$  and distance to  $\text{spt } \pi_n$  greater than  $k/2$  (for the  $\ell^1$  norm  $\|(x, y)\| = |x| + |y|$ . It is in fact close to  $k$ ). It follows

$$W^{\mathbb{R}^2}(\pi_n, \pi'_n) > \sum_{k=1}^{n-1} \frac{\varepsilon}{2(k+1)} = \frac{1}{2} \left( \sum_{k=2}^n \frac{1}{k} \right) \max(W(\mu_n, \mu'_n), W(v_n, v'_n)).$$

Note that we can normalise in mass and space and get the same estimate for families of probability measures close to  $\lambda_{[0,1]}$ . After this normalisation, the sequences  $\pi_n$  and  $\pi'_n$  both converge to  $(\text{Id} \otimes \text{Id})_{\#} \lambda_{[0,1]}$  but the ratios  $W^{\mathbb{R}^2}(\pi_n, \pi'_n) / W(\mu_n, \mu'_n)$  and  $W^{\mathbb{R}^2}(\pi_n, \pi'_n) / W(v_n, v'_n)$  go to infinity faster than  $\ln(n)/2$ . This makes it impossible to find an estimate like Corollary 2.32 for the Kantorovich distance in place of  $Z$ .

## References

- [1] M. Beiglböck, P. Henry-Labordère and F. Penkner. Model-independent bounds for option prices – A mass transport approach. *Finance Stoch.* **17** (3) (2013) 477–501. [MR3066985](#)
- [2] M. Beiglböck and N. Juillet. On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.* **44** (1) (2016) 42–106. [MR3456332](#)
- [3] P. Billingsley. *Convergence of Probability Measures*, 2nd edition. *Wiley Series in Probability and Statistics: Probability and Statistics*. Wiley, New York, 1999. [MR1700749](#)
- [4] A. Galichon, P. Henry-Labordère and N. Touzi. A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Ann. Appl. Probab.* **24** (1) (2014) 312–336. [MR3161649](#)
- [5] P. Henry-Labordère and N. Touzi. An explicit martingale version of Brenier’s theorem. *Finance Stoch.* **20** (3) (2016) 635–668. [MR3519164](#)
- [6] F. Hirsch and B. Roynette. A new proof of Kellerer’s theorem. *ESAIM Probab. Stat.* **16** (2012) 48–60. [MR2911021](#)
- [7] D. Hobson and A. Neuberger. Robust bounds for forward start options. *Math. Finance* **22** (1) (2012) 31–56. [MR2881879](#)
- [8] H. G. Kellerer. Markov-komposition und eine anwendung auf martingale. *Math. Ann.* **198** (1972) 99–122. [MR0356250](#)
- [9] T. Lindvall. On Strassen’s theorem on stochastic domination. *Electron. Commun. Probab.* **4** (1999) 51–59 (electronic). [MR1711599](#)
- [10] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Stat.* **36** (1965) 423–439. [MR0177430](#)
- [11] H. Thorisson. *Coupling, Stationarity, and Regeneration. Probability and Its Applications (New York)*. Springer, New York, 2000. [MR1741181](#)
- [12] C. Villani. *Topics in Optimal Transportation. Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI, 2003. [MR1964483](#)
- [13] C. Villani. *Optimal Transport. Grundlehren der Mathematischen Wissenschaften* **338**. Springer, Berlin, 2009. [MR2459454](#)